IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ryan

Serial No.: 10/608,463 Group Art Unit: 1652

Filed: June 27, 2003 Examiner: E. Slobodyansky

FOR: ISOLATED GENOMIC POLYNUCLEOTIDE FRAGMENTS FROM CHROMOSOME 12 THAT ENCODE HUMAN CARBOXYPEPTIDASE M AND THE HUMAN MOUSE DOUBLE MINUTE 2 HOMOLOG

Confirmation No.: 6428

SUPPLEMENTAL APPEAL BRIEF

Mail Stop Appeal
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I. REAL PARTY IN INTEREST

The real party in interest of the present application is Ryogen LLC . Ryogen LLC is the owner of the present application by way of an assignment from the inventor, James W. Ryan of all rights, title, and interest.

II. RELATED APPEALS AND INTERFERENCES

There are no appeals or interferences related to the present application.

III. STATUS OF CLAIMS

Claims 7, 10, 15-18, 20, 24, 25, 30 and 31 stand finally rejected by the Examiner as noted in the Advisory Action mailed October 29, 2007. Claims 1-6, 8-9, 11, 13, 19, 21 and

26-29 have been canceled. Claims 12, 14, 22, 23 and 32-38 have been withdrawn. The rejection of claims 7, 10, 15-18, 20, 24, 25, 30 and 31 are appealed.

IV. STATUS OF AMENDMENTS

No amendments have been submitted subsequent to the Examiner's Final Rejection.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

A. Independent Claim 7

Claim Elements	Support in specification
An isolated nucleic acid molecule 20-51039 contiguous nucleotides in length consisting of a reverse or forward strand of a region of SEQ ID NO:4	Page 9, line 34; page 10, lines 22-26
wherein said region is selected from the group consisting of a 5'-non coding region depicted in nucleotides 51039-41739 of SEQ ID NO:4, a 3'-non-coding region depicted in nucleotides 9503-1 of SEQ ID NO:4, a contiguous intron-exon region between nucleotides 41738-9502 of SEQ ID NO:4, wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2, a contiguous exon-intron region between nucleotide 41738-9502 of SEQ ID NO:4, wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2, an intron depicted in nucleotides 36385-40645, 36309-33127, 32994-29616, 29564-25577, 25507-25384, 25287-21169, 21006-14110, 13953-13267, and/or 13188-10665, a region comprising a dinucleotide of the following group: 41739-41738, 40645-40646, 36309-36310, 36384-36385, 32994-32995, 33126-33127, 29564-29565, 29615-29616, 25507-25508, 25287-25288, 25383-25384, 25576-25577, 21006-21007, 21168-21169, 14109-14110, 13953-13954, 13266-13267, 13188-13189, 10664-10665 and/or 9504-9503	Page 10: Table 2
a transcription binding site selected from the group consisting of BINDING SITES huMDM2, location in SEQ ID NO:4 AP1_C: 36-46, 2876-2886; AP4_Q5: 7944-7980; AP4_Q6: 7943-59, 8924-8940, 9294-9310;	Page 9, line 29 to page 10, line 2; Table 3 on pages 10-12

ARNT_01: 1682-1706, 2193-2217, 9201-9225;

BRN2_01: 1040-1058, 7803-7821;

CAAT_01: 3292-3306;

CDPCR3HD_01: 6522-6540;

CEBPB_01: 1424-1438, 3917-3931, 4178-4192, 4787-4801, 6855-6869;

CREL_01: 5630-5642;

DELTAEF1_01: 83-95, 6328-6340;

FREAC7_01: 2757-2773, 5154-5170, 5823-5839;

GATA1_04: 4846-4858, 7017-7029;

GATA1_05: 8464-8476;

GATA2_02: 6045-6057, 6073-6085, 6142-6154;

GATA2_03: 2489-2501, 3323-3335, 3384-3396, 7393-7405:

GATA3_02: 3264-3276, 6870-6882:

GATA3_03: 40-52, 5729-5741, 6529-6541, 6874-6886, 7041-7053, 7589-7601;

GATA_C: 7 349-7361, 8188-8200;

HFH2_01: 1743-1759, 7995-8011;

HFH3_01: 502-518, 1739-1755, 4160-4176, 9402-9418, 9418-9434;

HFH8_01: 8184-8200;

IK2_01: 951-963, 3588-3600;

MZF1_01: 1202-1210, 1447-1455, 4997-4005, 5424-5432;

NF1_Q6: 1480-1500, 8166-8182;

NFAT_Q6: 4190-4208, 6009-6027;

NKX25_01: 741-755, 1648-1662, 1885-1899, 1984-1998, 3609-3623, 4928-4942, 5060-5074, 5889-5903, 8850-8864, 9190-9204;

NKX25_02: 2584-2599, 2970-2984, 4644-4658, 5179-5193, 6482-6496;

NMYC_01: 2560-2572;

RORA1_01: 220-238, 2638-2656;

S8_01: 4644-4656, 4842-4854, 4845-4857, 5200-5212, 5371-5383, 5735-5747, 6482-6494, 6541-6553, 6544-6556, 6772-6784, 7270-7292, 7273-7285;

SOX5_01: 1355-1371, 1430-1446, 3094-3110, 3155-3171, 4669-4685, 4692-4708, 4789-4805;

SRY_02: 4164-4180, 5665-5681;

TATA_01: 1261-1277, 2574-2590, 2723-2739, 2733-2749, 2770-2786, 4199-4215, 4206-4222;

TATA_C: 5900-5916, 7456-7472, 7702-7718, 7917-7933; and

XFD2_01: 7702-7218, 7917-7933;

a transcription binding site selected from the group consisting of

BINDING SITES huMDM2, location in SEQ ID NO:4
AP1_C: 12109-12119, 12695-12705, 22600-22610, 24166-24176, 31311-31321, 35234-35244, 39184-39194;

AP1_Q2: 11952-11962, 12068-12078, 14798-14808, 21748-21758, 22613-22623, 23676-23686, 26562-26572, 30046-30056;

AP1_Q4: 12695-12705, 31311-31321, 35234-35244, 36295-36305, 38784-38794, 39188-39198;

AP4_Q6: 31635-31651;

BRN2_01: 13448-13466, 14764-14782, 28094-28112, 40027-40045; CAAT_01: 11288-11302, 15054-15068; CDPCR3HD_01: 11286-11304, 13284-13302, 20846-20864, 29344-29362; CEBPB_01: 29241-29255; CREL_01: 36091-36103, 38873-38885; DELTAEF1_01: 18083-18095, 20385-20397, 26955-26967; FREAC7_01: 11982-11998, 15187-15202, 16523-16539, 16529-16545, 16587-16603, 16604-16620, 16676-16642, 16633-16649, 16644-16660, 16650-16666, 16657-16673, 16673-16689, 16762-16778, 21332-21348, 25689-25700, 26529-26545, 27767-27783, 29495-29511; GATA1_02: 10916-10928, 15775-15789, 18162-18174, 26088-26100, 32518-32530; GATA1_03: 28012-28024; GATA1_04: 11153-11165, 11630-11642, 13778-13790, 17439-17451, 19300-19312, 21606-21618, 22743-22755, 23747-23759, 25806-25818, 26529-26541, 29424-29436, 30455-30467, 32761-32778, 33352-33364, 33960-33972, 36101-36113, 40007-40019; GATA1_05: 11590-11602, 26550-26562, 36737-36749; GATA1_06: 18772-18784, 23054-23066, 35568-35580, 37855-37867; GATA2_02: 20755-20767, 30830-30842, 34755-34767, 36285-36297, 39143-39155, 39641-39653, 40586-40598; GATA2_03: 13535-13547, 22711-22723, 23161-23173, 25028-25040, 27237-27249, 36277-36289; GATA3_02: 11558-11570, 16470-16482, 17225-17237, 19619- 19631, 22156-22168, 22443-22455, 24713-24725, 27619-27631, 32716-32728, 34124-34136, 34163-34175, 36832-36844,	

38403-38415; GATA3_03: 10869-10881, 11515-11527, 13845-13857, 17221-17233, 18952-18964, 20050-20062, 40171-40183; GATA_C: 15848-15860, 18899-18911, 23640-23652, 29072-29084, 30881-30893, 33198-33210, 37472-37484, 38621-38633; GFI1_01: 35469-35481, 35492-35504; HFH2_01: 15939-15955, 24636-24652, 25866-25882, 32171-32187, 35372-35388, 39457-35473; HFH3_01: 13340-13356, 19218-19234, 21328-21344, 21336-21352, 21344-21360, 28062-28078, 32125-32141; HFH8_01: 14133-14149, 22578-22584; HNF3B_01: 13150-13166, 16505-16521, 25264-25280, 29443-29459, 37654-37670; IK2_01: 11547-11559, 17144-17156, 18961-18973, 23883-23895, 27617-27629, 28908-28920, 29241-29253, 30752-30764, 34768-34780; LYF1_01: 12319-12331, 19191-19203, 37226-37238, 39430-39442; MAX_01: 22974-22986, 33339-33351; MZF1_01: 26105-26113, 35187-35195; NF1_Q6: 12048-12064, 33334-33354; NFAT_Q6: 13295-13313, 14157-14175, 14311-14329, 14414-14432, 18269-18287. 19326-19344, 20801-20819, 21177-21195, 22537-22555, 23861-23879, 25392-25410, 25879-25897, 27524-27542, 30636-30654, 30718-30736, 31525-31543, 33655-33673, 34726-34744, 34917-34535, 34990-35008, 35979-35997, 36479-36493, 36577-36595, 37154-37172, 40224-40242, 40365-40383; NKX25_01: 12041-12055, 12340-12354, 12471-12485, 12742-12756, 12877-12891, 13849-13863, 18995-19009, 21440-21454, 21883-21897, 28426-28440, 30964-30978, 32033-32047, 32265-	

32279; NKX25_02: 10998-11012, 12711-12725, 14131-14145, 14726-14740, 16024-16038; NMYC_01: 18753-18765, 18754-18766, 23076-23088, 30534-30546, 34400-34412; RORA1_01: 13134-13152, 22966-22984, 24934-24952, 33341-33359, 34760-34778; S8_01: 11000-11012, 11977-11989, 12048-12060, 12051-12063, 13747-13759, 13923-13935, 13926-13938, 14676-14688, 14679-14691, 16026-16038, 16313-16325, 16316-16328, 17515-17527, 20756-20768, 20759-20771, 23154-23166, 23157-23169, 25198-25210, 25201-25213, 26651-26663, 27508-27520, 27511-27523, 29450-29462, 29478-28490, 29775-29787, 29778-29790, 29813-29825, 29816-29828, 31329-31341, 31677-31689, 31680-31692, 31732-31744, 31735-31747, 36137-36149, 36140-36152, 36812-36824, 36815-36827, 37413-37425, 38679-38691, 39474-39486, 39477-39489; SOX5_01: 27397-27413, 27572-27588, 28100-28116, 29230-29246, 29439-29455, 30690-30706, 31595-31611, 33871-33887, 34113-34129, 34624-34640, 37668-37684, 38582-38598, 39124-39140, 40410-40426; SRY_02: 20016-20032, 22410-22426, 27329-27345, 29162-29178, 29499-29515, 30646-30662, 31503-31519, 35928-35944, 37324-37340; TATA_01: 32722-32738, 32729-32745, 32807-32823, 33825-33841, 34120-34136, 35433-35449, 36593-36609; TATA_C: 11015-11031, 11817-11833, 13635-13651, 14930-14946; TCF11_01: 18543-18549, 22574-22580, 31281-31297, 31489-31505, 38754-38770; USF_01: 23075-23087, 32577-32589; VMYB_02: 11526-11538, 17384-17396, 18400-18412, 19549-19561, 22188-22200, 40486-40508 and XFD2_01: 16620-16636. 18153-18169, 22102-22118, 23141-	

23157. And a transcription binding site selected from the group consisting of BINDING SITES huMDM2 location in SEQ ID NO:4 AP1_C: 44584-44594, 49069-49079: AP1_Q2: 42174-42184, 45217-45227, 48422-48422, 50447-50457; AP1_Q4: 42702-42712, 50806-50816; AP4_Q6: 42117-42133, 42118-42134, 42244-42260, 45432-45448; 45433-45449, 46609-46625; BRN2_01: 42310-42328, 44022-44040, 47514-47532, 48900-48918, 48967-48985; CAAT_01: 44866-44880; CDPCR3HD_01: 45671-45689, 49219-49237; CREL_01: 42437-42449, 49797-49809; FREAC7_01: 47026-47042, 47292-47308, 47658-47674; GATA1_02: 43482-43494, 48926-48938, 49284-49296; GATA1_03: 47371-47383; GATA1_04: 43054-43066, 43162-43162, 43967-43979, 45464-45476, 45916-45928, 47763-47775; GATA1_05: 49319-49331, 49459-49471; GATA1_06: 47590-47602; GATA2_02: 42660-42672, 43475-43487; GATA2_03: 43714-43726, 50948-50960; GATA3_02: 49155-49167, 49844-49856;	

GATA3_03: 42202-42214, 44810-44822, 48438-48450, 49136-49148, 49337-49349, 49869-49881; GATA_C: 44011-44023, 45256-45268, 45823-45835, 47915-47927, 49201-49213, 49573-49585; GFI1_01: 46606-46618, 47063-47075; HFH3_01: 47030-47046, 47284-47300, 47288-47304; IK2_01: 45275-45287; LYF1_01: 44564-44576, 46991-47003, 49567-49579; MAX_01: 43234-43246, 48726-48738; MZF1_01: 41772-41780, 42290-42298, 42295-42303, 44507-44515, 45105-45113, 45203-45211, 49948-49956, 50774-50782; NF1_Q6: 50209-50229; NFAT_Q6: 42061-42079, 44418-44436, 46399-46417, 47974-47992, 49267-49285, 49964-49982, 50392-50410; NKX25_01: 42394-42408, 43507-43521, 46115-46129; RORA1_01: 45073-45091, 48718-48736; S8_01: 43552-43564, 45214-45226, 47160-47172, 48419-48431, 49295-49307, 50379-50391; SOX5_01: 43716-43732, 46351-46367, 47156-47172, 47774-47790, 47868-47884, 47974-47990, 48915-48931, 50323-50339; TATA_01: 45588-45604, 47625-47641, 48026-48042, 48659-48675, 49056-49072, 49079-49095, 49152-49168; TCF11_01: 49115-49131; VMYB_02: 42010-42022, 42279-42291, 44651-44663; XFD2_01: 42870-42886, 42910-42926.	

B. Independent Claim 24

Claim Element	Support in Specification
An isolated nucleic acid molecule 20-5000 contiguous nucleotides in length	Page 9, line 35 to page 10, line 2
consisting of a reverse or forward strand of a contiguous exon intron region between nucleotides 41738-9502 of SEQ ID NO:4 or contiguous intron-exon region between nucleotides 41738-9502 of SEQ ID NO:4,	Page 10, Table 2 and page 14, lines 29-32
wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2	Page 2, lines 7-13; page 10, Table 2

VI. Grounds of Rejection to Be Reviewed on Appeal

Whether claim 7, 10, 15-18, 20, 24, 25, 30 and 31 are unpatentable over Muzny et al. ("Muzny") in view of Vogelstein et al. ("Vogelstein").

VI. Argument

In the final rejection dated April 16, 2007, the Office Action stated:

> It would have been obvious to one of ordinary skill in the art at the time the invention was made to use said cDNA to identify the genomic DNA that encodes the human MDM2 homolog of SEQ ID NO:2 on chromosome 12q12-14. The motivation is provided by Vogelstein et al. who teach that it binds to oncogene p53 and is diagnostic of tumorigenesis. The state of the art provides various techniques for obtaining genomic DNA using cDNA probes that are usually labeled. The comparison of genomic and cDNA would result in the identification of regions comprising exon-intron and intron-exon junctions within coding and non-coding regions. One of ordinary skill in the art would have been motivated to use said non-coding regions or fragments thereof of at least 20 nucleotides and up to 5000 or 51039 nucleotides (the entire length of SEQ ID NO:4) nucleotides for detecting splice variants of chromosome 12q12-14 from genomic nucleotide samples from an individual, for example. As a matter of convenience a non-coding region such as an exon-intron or intron-exon region or fragments thereof can be present in a kit or on a solid support. Further, said support can be a microarray according to a customary use of nucleic acid molecules in the art.

Appellant respectfully traverses the rejection. In Appellant's view, it would not have been obvious to combine the disclosure of Muzny and Vogelstein given that there was no suggestion to do so. Muzny merely contains just a small portion of chromosome 12 DNA. Chromosome 12 is about 130 million base pairs long and is believed to contain several hundred genes (by analysis after 2001 and after the Applicant discovered the human MDM2 homologue gene). Muzny et al knew that clone AC025423 (from 1V11-61102) was from chromosome 12 but there is no evidence in the NCBI report of a sub-assignment to the p- or q-arm. Further, there is no evidence that Muzny knew whether the clone did or did not contain one or more genes and particularly whether it contained the gene encoded by SEQ ID NO:4. As will be discussed in further detail below, the MDM2 cDNA constitutes just 1.6% of the clone disclosed by Muzny. Undue experimentation would have been required not only to locate the MDM2 gene but also identify exon-intron junctions.

Even assuming *arguendo* that there had been such a motivation, one of ordinary skill in the art would not have obtained the isolated nucleic acid molecules of the present invention, the specifically recited non-coding regions of the MDM2 gene located between nucleotides 20-51039 of SEQ ID NO:4. This is because, as noted in previous responses, Vogelstein placed the human MDM2 homologue gene at 12q12-14. Actually, this finding is incorrect. After the publication of Vogelstein, the gene was found not to be located at 12q12-14. The gene is actually several millions of base pairs away at 12q15 (see, for example, Andersen et al., 1996, Mammalian Genome 7:780-783, Bureau, 1995, Genomics 28: 109-112 and Genecard, attached hereto as Exhibit 1 and previously made of record). Thus, even if these two references were indeed combined, the ordinary skilled artisan would have looked for the MDM2 gene in the wrong location and thus would not have obtained the claimed sequences.

In response to Remarks made by Appellant in the amendment submitted on January 23, 2007, the final Office Action states

> With regard to the 103(a) rejection, Applicant argues that "There was certainly no indication given in the cited art either singly or in combination regarding the location of the MDM2 gene encoding human mouse double minute 2 homolog depicted in SEQ ID NO:2 on AC025423" (Remarks, page 13). As was explained in the previous Office action mailed August 25, 2006 "the exact location of the gene is not necessary as long

as its sequence is known as in the instant case" (page 5). At the time the invention was made finding non-coding regions using cDNA and genomic DNA was standard technique. Watson et al (1992) "Recombinant DNA" teach that once the first genes were cloned, introns were identified by comparing the cloned genomic DNA with the corresponding cloned cDNA (page 137, 2nd column). In the case of the instant application, both genomic and cDNA were known. They only needed to be compared in order to identify intron-exon junctions. Applicant further argues that "There is no prior art that defines the complete genomic structure of a particular gene. This is necessary in order to identify the claimed noncoding sequences in the instant invention" (page 13). This argument is similar to the issue of location and is responded above. Applicants further argues that "One of ordinary skill in the art would have no idea as to the number of introns and the length of the 5' and 3' noncoding sequences in the MDM2 gene" (page 14). It is agreed that said number and length were not known before the invention. If they were known, the rejection would be 102. The current rejection is 103(a) stating that it would have been obvious to compare genomic DNA and cDNA and identify the number of introns and the length of the 5' and 3' noncoding sequences in the MDM2 gene. Applicant further argues "The Examiner's assertion that one of ordinary skill in the art would have expected that the location is often imprecise actually further supports Applicant's assertion the claimed sequences were indeed nonobvious. If the location is imprecise, where would one of ordinary skill in the art know where to look?" (page 15). This is not persuasive because the precise location is not necessary when the 2 sequences that need to be compared are known. They would be obvious because the genomic DNA was already sequenced and cDNA was made. Applicant further argues that "It should be noted that annotation of the human genomic DNA was still relatively new as of the priority date of the instant application. Even assuming arguendo that finding noncoding regions using cDNA and genomic DNA was standard technique, the means to make the invention does not predict the claimed invention. Specifically the means used to make the invention do not predict the claimed nucleic acid molecules. BLASTN, TBLASTN, etc. do not themselves predict gene-specific results. It is Applicant's view that only general guidance is provided. This is not sufficient" (page 16). This is not persuasive because there is no need to predict the sequence itself, it was known before. The invention is in the identification of the specific fragments (exons/introns) of the known sequence. Applicant appears to argue as if the genomic

> DNA was not sequenced prior to the instant invention. Applicant further argues that "it is Applicant's view that given that the cDNA constitutes just 1.6% of the AC025423 sequence is in itself evidence of the unpredictability of determining the entire sequence of the MDM2 gene and thus contiguous intron-exon and exon-intron regions. The Examiner is in effect asserting that just because Applicant did isolate the claimed nucleic acid molecule, it must have been obvious to do so. It is well established case law that the fact that the inventors were ultimately successful is irrelevant to whether one of ordinary skill in the art at the time the invention was made would have reasonably expected success" (paragraph bridging pages 16-17). This is not persuasive because the size of the cDNA does not matter. ESTs of smaller size are used for comparison to genomic sequences. Applicant does not show what unexpected difficulties other than routine comparison of the genomic DNA and cDNA were encountered during the time the invention was made.

It appears that the Examiner is asserting that just because a genomic clone containing chromosome 12 sequences had been isolated, the MDM2 cDNA was known and that techniques were known for finding noncoding regions, one of ordinary skill in the art would have had a reasonable expectation of success of obtaining the claimed sequences. Further, the Examiner is asserting that it is not significant that cDNA just constitutes just 1.6% of the AC025423 sequence. Appellant disagrees for a number of reasons.

First, Appellant asserts that the Watson disclosure is of limited relevance here. Specifically, Watson specifically states on page 137

> It should be noted that at the time that the electron microscopy experiments on adenovirus were done, no one had clone a cellular gene yet. Once the first genes were cloned, introns were identified by comparing the cloned genomic DNA with the corresponding cloned cDNA.

This is very different from the situation with respect to the isolated non-coding fragments of the instant invention. Muzny merely discloses the sequence of a genomic clone containing chromosome 12 sequences, not isolated SEQ ID NO:4. No indication is provided is provided in the Muzny disclosure as to which portion of chromosome 12 has actually been sequenced. Vogelstein merely discloses the MDM2 cDNA. Applicant was the first to identify SEQ ID NO:4 and determine that it did indeed encode MDM2 and in particular determine the claimed noncoding sequences.

Second, Appellant asserts that there would not be a reasonable expectation of success of obtaining the claimed noncoding sequences of SEQ ID NO:4 in view of the cited references. Vogelstein placed the human MDM2 homologue gene at 12q12-14. As noted above, there was actually a previous disclosure stating that the MDM2 was located between 12q14.3-15 (see, for example, Andersen et al., 1996, Mammalian Genome 7:780-783 and Bureau, 1995, Genomics 28: 109-112, submitted and disclosed in previous response attached hereto as Exhibit 1). However, given the conflicting locations published as of the priority date of the instant application, one of ordinary skill in the art would not have known which location was actually correct. This situation in Appellant's view would constitute undue experimentation.

Third, Appellant disagrees with the assertion that the size of the cDNA does not matter since ESTs of smaller size are used for comparison to genomic sequences. In Appellant's view, the size of the cDNA relative to the isolated genomic clone is of particular significance. If, for example, the cDNA constituted 50% of a particular genomic clone, considerable less experimentation would be involved in determining the sequence of a particular gene than when it is merely 1.6% of the genomic clone.

B. Advisory Action

The Advisory Action states:

> Applicant reiterates the previous arguments that have been answered in the Final Office Action mailed 4/16/07. As correctly noted by Applicant, the examiner asserts that if genomic and cDNA sequences are known, one of ordinary skill in the art would have a reasonable expectation of success of obtaining the claimed sequences (Remarks, page 15). Applicant does not explain why the fact that cDNA constitutes 1.6% of the genomic sequence leads to no expectation of success. Applicant further discusses KSR v Teleflex case, stating that "under KSR, an approach that is obvious to try is also obvious where normal trial and error procedures will lead to the result" (page 16). It appears that KSR supports the 103(a) rejection because Applicant does not show why the result wouldn't be obtained.

In response, it has throughout prosecution been asserted that the fact that the MDM2 cDNA constitutes just 1.6% of the genomic sequence disclosed by Muzny is very significant

since in Appellant's view, undue experimentation would be involved to isolate the claimed sequence. More than trial and error would be involved.

Muzny only disclosed the sequence of the clone AC025423. There was no indication at the time of the filing date of the instant application that AC025423 actually contained SEQ ID NO:4, more specifically a sequence encoding human MDM2. Muzny did not sequence all of chromosome 12; they only sequenced the chromosome 12 BAC clone AC025423, less than 0.12% of the chromosome (about 157 kB vs. about 132,000 Kb). The clone's sub-site on chromosome 12 apparently was not then known. Muzny did not even suggest, must less state whether the AC025423 did or did not contain any gene. The determination of a sequence of genomic DNA does not necessarily imply the presence of a gene. This is because it is well known in the art and was certainly well known when the instant invention was made and as of the filing date of the instant application that most genomic DNA contains "junk" DNA, not genes. Appellant attaches hereto as Exhibit 2 the following review article (Wong et al., 2000, "Is Junk DNA Mostly Intron DNA?" Genome Research 10: 1672-1678) which discuss "junk" DNA in further detail. As is evident from this review article, junk DNA was well known as of the priority date of the above-referenced application. When obtaining genomic clone sequences, one of skill in the art would not have any way of knowing whether or not it actually contains a gene(s) or just junk DNA. Even noting that a clone has a high GC content is a poor guide to a sequence's likely gene content, especially given the presence of pseudogenes. Therefore, *contra* to assertions made in the Office Action, disclosure of this particular sequence would not indicate to the skilled artisan that this clone would necessarily contain a gene.

The clone AC025423 is 150,579 nucleotides in length. The cDNA sequence only contains 2372 nucleotides (1.6% of AC025423). However, one of ordinary skill in the art would not know where or how these 2372 nucleotides are interspersed within the AC025423 clone or if indeed it is even present. No teachings are provided as to the structure of the MDM2 gene itself: number and size of exons, number and size of introns, locations of exons and introns and number and size of 5' and 3' untranslated regions. The possibilities are close to infinite. Thus there would not be a reasonable expectation of success of obtaining the claimed sequences given that the isolation and identification of the claimed sequences constitutes undue experimentation.

It is Appellant's view that the combination of the two references would still constitute "obvious to try" even under KSR v Teleflex. *KSR v Teleflex*, 82 USPQ2d 1385; 127 S.Ct. 1727 (2007). Appellant notes that the "Examination Guidelines for Determining Obviousness Under 35 U.S.C. 103 in View of the Supreme Court Decision in *KSR International Co. v. Teleflex*", 72 FR 57526 (October 10, 2007) (hereinafter "Examination Guidelines") stated with respect to "obvious to try" the following:

> To reject a claim based on this rationale, Office personnel must resolve the *Graham* factual inquiries. Office personnel must then articulate the following:
>
> (1) a finding that at the time of the invention, there had been a recognized problem or need in the art, which may include a design need or market pressure to solve a problem;
>
> (2) a finding that there had been a finite number of identified, predictable potential solutions to the recognized need or problem;
>
> (3) a finding that one of ordinary skill in the art could have pursued the known potential solutions with a reasonable expectation of success; and
>
> (4) whatever additional findings based on the *Graham* factual inquiries may be necessary, in view of the facts of the case under consideration, to explain a conclusion of obviousness.

Ex parte Kubin, 83 USPQ2d 1410, 2007 WL 2070495 (Bd. App. & Int. 2007) was cited as an example in the Examination Guidelines of a situation where a finite number of identified, predictable solutions are provided with a reasonable expectation of success. In *Kubin*, the claimed invention was directed to

> An isolated nucleic acid molecule comprising a polynucleotide encoding a polypeptide at least 80% identical to amino acids 22-221 of SEQ ID NO:2, wherein the polypeptide binds CD48.

Id at 1412.

In the rejection affirmed by the Board, the Examiner asserted that

> The skilled artisan would have been motivated to isolate the nucleic acid sequence corresponding to NAIL, based on Valiante's disclosure of p38 (which is the same protein as NAIL) and Valiante's express teachings how to isolate p38 cDNA by using conventional techniques, such as taught in Sambrook, including using mAbC1.7, a probe specific for p38.

Id at 1412.

This finding of obviousness in *Kubin* appeared to be predicated on the obviousness of isolating NAIL cDNA, not any other nucleic acid sequences encoding the NAIL protein. The Board specifically states

> Based on our findings and those of the Examiner, at least one of Appellants' claimed polynucleotides would have been obvious to one of ordinary skill in the art at the time Appellants' invention was made.....
>
> Appellants argue the "cited references do not provide an adequate written description of the claimed nucleic acid sequences." (citation omitted). In so arguing, Appellants overlook the distinction between obviousness under §103 and lack of written description under §112, § 1. A single obvious species within a claimed genus renders the claimed genus unpatentable under §103. Thus, an obvious method of obtaining a single nucleic acid molecule encoding NAIL may be all that is required to show that the presently claimed genus of nucleic acid molecules is unpatentable under §103.....
>
> The "problem" facing those in the art was to isolate NAIL cDNA, and there were a limited number of methodologies available to do so. The skilled artisan would have had reason to try these methodologies with the reasonable expectation that at least one would be successful.

Id at 1414.

The claimed invention is directed to a nucleic acid molecule 20-51039 contiguous nucleotides in length consisting of a reverse or forward strand of a region of SEQ ID NO:4. SEQ ID NO:4 is a genomic sequence, not cDNA encoding MDM2. This is very different from *Kubin* which involves deducing the NAIL cDNA sequence from NAIL polypeptide sequence and subsequently isolating this sequence. The art cited in the instant application was a cDNA sequence and a genomic clone containing chromosome 12 sequence. Considerably more and undue experimentation would be involved in identifying and isolating the genomic sequence and consequently the claimed regions comprising the recited noncoding regions of the MDM2 gene than in identifying cDNA in view of a disclosure of a polypeptide sequence. The claimed sequence is between 20-51039 nucleotides in length. The Muzny sequence is 150,579 nucleotides in length. No direction or frame of reference is provided as to where the MDM2 sequence could be located. Furthermore, although Vogelstein, disclosed the MDM2

cDNA, no direction is provided as to the location of exons, introns, 5' untranslated and/or 3' untranslated regions. The MDM2 cDNA was only 2372 nucleotides in length. Thus, there would have been a huge number of strategies and methodologies that one of ordinary skill in the art could have employed in attempting to obtain the claimed sequences.

Appellant also notes that in the Examination Guidelines it is stated

> The key to supporting any rejection under 35 U.S.C. 103 is the clear articulation of the reason(s) why the claimed invention would have been obvious. The Supreme Court in KSR noted that the analysis supporting a rejection under 35 U.S.C. 103 should be made explicit. The Court quoting *In re Kahn* stated that ""'[R]ejections on obviousness cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.'"

This statement follows the holding of established case law, most notably *In re Beasley* 117 Fed. Appx. 739, 2004 WL 2793170 (C.A. Fed 2004). In *Beasley,* the Court held that:

> The examiner and the Board have managed to find motivation for substituting one type of memory for another without providing a citation of any relevant, identifiable source of information justifying such substitution. The statements made by the Examiner, upon which the Board relied, amount to no more than conclusory statements of generalized advantages and convenient assumptions about skilled artisans.

A similar situation exists with respect to the instant application. An example of such conclusory statements made are on page 4 of the Office Action and shown in boldface:

> It would have been obvious to one of ordinary skill in the art at the time the invention was made to use said cDNA to identify the genomic DNA that encodes the human MDM2 homolog of SEQ ID NO:2 on chromosome 12q12-14. The motivation is provide by Vogelstein et al. who teach that it binds to oncogene p53 and is diagnostic of tumorigenesis. **The state of the art provides various techniques for obtaining genomic DNA using cDNA probes that are usually labeled. The comparison of genomic and cDNA would result in the identification of regions comprising exon-intron and intron-exon junctions within coding and non-coding regions.**

The Examiner has only referred to Watson, 1992. As previously argued, Watson would not apply in this case since in Watson, the genes themselves were actually cloned. No

other references were provided by the Examiner in response to Appellant's subsequent arguments. Furthermore, there is nothing stated in the Office Action as to how the state of the art teaches how given the teaching of a large genomic clone and the cDNA sequence of a particular gene, one of ordinary skill in the art could with particularity identify specific exon and intron sequences of a particular gene and assemble it in its entirety. There is no prior art that defines the complete genomic structure of a particular gene. This is necessary in order to accurately identify the claimed noncoding sequences in the instant invention.

C. Secondary Considerations

Appellant also asserts that secondary considerations should also apply when determining if the claimed invention is obvious over Muzny in view of Vogelstein. One secondary consideration of particular relevance is long felt need. Specifically, there has been a great deal of interest in the scientific community in MDM2 given its potential use as a diagnostic and therapeutic agent. This interest is summarized in the cited patent, Vogelstein Additionally, Appellant submitted during prosecution and IDS listing four references relating to MDM2 (a copy of the 1449 Form previously submitted and made of record is attached hereto as Exhibit 3). However, there was absolutely no disclosure or suggestion of the genomic organization of MDM2 genomic DNA until the instant application was filed. An independent disclosure of the genomic organization of the MDM2 gene was not available until July 21, 2004, more than one year after the filing date of the instant application (Liang et al., 2004, Gene 338:217-223, submitted herewith as Exhibit 4). The dearth of knowledge regarding the MDM2 gene was actually admitted to by Liang et al. on the first page of his article where it was stated "Although the human MDM2 cDNA sequence has been reported, the genomic organization of the human gene has not been documented". Further, Liang et al. discussed the usefulness of determining the genomic structure and organization of the MDM2 gene on page 218:

> One of the distinctive properties of MDM2 is the possession of an extremely complex expression pattern. Its multiple-sized transcripts and proteins have been found in tumour samples and cell lines by a number of groups (citations omitted). In our previous studies, five alternatively sized transcripts of the human MDM2 were found in human ovarian tumour, bladder tumour and leukaemic cell samples (citation omitted).....Here

> we present data demonstrating two further MDM2 transcript forms with internal sequence deletions in human tumour tissue. We Hypothesised that these transcripts are generated by alternative splicing. To test this hypothesis and to explore the associated mechanisms, we have investigated the genomic structure and organization of the human MDM2 gene…

Conclusion

It is Appellant's position that claims are not obvious over Muzny in view of Vogelstein. It is further Appellant's position that the claims are in condition for allowance. The Examiner is invited to contact the undersigned at (914) 712-0093 if she has any questions.

Respectfully submitted,

Date: March 2, 2008

/Cheryl H Agris/

Cheryl H. Agris, Reg. No. 34,086
P.O. Box 806
Pelham, N.Y. 10803
(914) 712-0093
Customer No. 25538

CLAIMS APPENDIX

7. An isolated nucleic acid molecule 20-51039 contiguous nucleotides in length consisting of a reverse or forward strand of a region of SEQ ID NO:4, wherein said region is selected from the group consisting of a 5'-non coding region depicted in nucleotides 51039-41739 of SEQ ID NO:4, a 3'-non-coding region depicted in nucleotides 9503-1 of SEQ ID NO:4, a contiguous intron-exon region between nucleotides 41738-9502 of SEQ ID NO:4, wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2, a contiguous exon-intron region between nucleotide 41738-9502 of SEQ ID NO:4, wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2, an intron depicted in nucleotides 36385-40645, 36309-33127, 32994-29616, 29564-25577, 25507-25384, 25287-21169, 21006-14110, 13953-13267, and/or 13188-10665, a region comprising a dinucleotide of the following group: 41739-41738, 40645-40646, 36309-36310, 36384-36385, 32994-32995, 33126-33127, 29564-29565, 29615-29616, 25507-25508, 25287-25288, 25383-25384, 25576-25577, 21006-21007, 21168-21169, 14109-14110, 13953-13954, 13266-13267, 13188-13189, 10664-10665 and/or 9504-9503; a transcription binding site selected from the group consisting of

BINDING SITES	huMDM2, location in SEQ ID NO:4
AP1_C:	36-46, 2876-2886;
AP4_Q5:	7944-7980;

AP4_Q6:	7943-59, 8924-8940, 9294-9310;
ARNT_01:	1682-1706, 2193-2217, 9201-9225;
BRN2_01:	1040-1058, 7803-7821;
CAAT_01:	3292-3306;
CDPCR3HD_01:	6522-6540;
CEBPB_01:	1424-1438, 3917-3931, 4178-4192, 4787-4801, 6855-6869;
CREL_01:	5630-5642;
DELTAEF1_01:	83-95, 6328-6340;
FREAC7_01:	2757-2773, 5154-5170, 5823-5839;
GATA1_04:	4846-4858, 7017-7029;
GATA1_05:	8464-8476;
GATA2_02:	6045-6057, 6073-6085, 6142-6154;
GATA2_03:	2489-2501, 3323-3335, 3384-3396, 7393-7405:

GATA3_02: 3264-3276, 6870-6882:

GATA3_03: 40-52, 5729-5741, 6529-6541, 6874-6886, 7041-7053, 7589-7601;

GATA_C: 7 349-7361, 8188-8200;

HFH2_01: 1743-1759, 7995-8011;

HFH3_01: 502-518, 1739-1755, 4160-4176, 9402-9418, 9418-9434;

HFH8_01: 8184-8200;

IK2_01: 951-963, 3588-3600;

MZF1_01: 1202-1210, 1447-1455, 4997-4005, 5424-5432;

NF1_Q6: 1480-1500, 8166-8182;

NFAT_Q6: 4190-4208, 6009-6027;

NKX25_01: 741-755, 1648-1662, 1885-1899, 1984-1998, 3609-3623, 4928-4942, 5060-5074, 5889-5903, 8850-8864, 9190-9204;

NKX25_02: 2584-2599, 2970-2984, 4644-4658, 5179-5193, 6482-6496;

NMYC_01: 2560-2572;

RORA1_01: 220-238, 2638-2656;

S8_01: 4644-4656, 4842-4854, 4845-4857, 5200-5212, 5371-5383, 5735-5747, 6482-6494, 6541-6553, 6544-6556, 6772-6784, 7270-7292, 7273-7285;

SOX5_01: 1355-1371, 1430-1446, 3094-3110, 3155-3171, 4669-4685, 4692-4708, 4789-4805;

SRY_02: 4164-4180, 5665-5681;

TATA_01: 1261-1277, 2574-2590, 2723-2739, 2733-2749, 2770-2786, 4199-4215, 4206-4222;

TATA_C: 5900-5916, 7456-7472, 7702-7718, 7917-7933; and

XFD2_01: 7702-7218, 7917-7933; .

a transcription binding site selected from the group consisting of

BINDING SITES huMDM2, location in SEQ ID NO:4

AP1_C: 12109-12119, 12695-12705, 22600-22610, 24166-24176, 31311-31321, 35234-35244, 39184-39194;

AP1_Q2: 11952-11962, 12068-12078, 14798-14808, 21748-21758, 22613-22623, 23676-23686, 26562-26572, 30046-30056;

AP1_Q4: 12695-12705, 31311-31321, 35234-35244, 36295-36305, 38784-38794, 39188-39198;

AP4_Q6: 31635-31651;

BRN2_01: 13448-13466, 14764-14782, 28094-28112, 40027-40045;

CAAT_01: 11288-11302, 15054-15068;

CDPCR3HD_01: 11286-11304, 13284-13302, 20846-20864, 29344-29362;

CEBPB_01: 29241-29255;

CREL_01: 36091-36103, 38873-38885;

DELTAEF1_01: 18083-18095, 20385-20397, 26955-26967;

FREAC7_01: 11982-11998, 15187-15202, 16523-16539, 16529-16545, 16587-16603, 16604-16620, 16676-16642, 16633-16649, 16644-16660, 16650-16666, 16657-16673, 16673-16689, 16762-16778, 21332-21348, 25689-25700, 26529-26545, 27767-27783, 29495-29511;

GATA1_02: 10916-10928, 15775-15789, 18162-18174, 26088-26100, 32518-32530;

GATA1_03: 28012-28024;

GATA1_04: 11153-11165, 11630-11642, 13778-13790, 17439-17451, 19300-19312, 21606-21618, 22743-22755, 23747-23759, 25806-25818, 26529-26541, 29424-29436, 30455-30467, 32761-32778, 33352-33364, 33960-33972, 36101-36113, 40007-40019;

GATA1_05: 11590-11602, 26550-26562, 36737-36749;

GATA1_06: 18772-18784, 23054-23066, 35568-35580, 37855-37867;

GATA2_02: 20755-20767, 30830-30842, 34755-34767, 36285-36297, 39143-39155, 39641-39653, 40586-40598;

GATA2_03: 13535-13547, 22711-22723, 23161-23173, 25028-25040, 27237-27249, 36277-36289;

GATA3_02: 11558-11570, 16470-16482, 17225-17237, 19619- 19631, 22156-22168, 22443-22455, 24713-24725, 27619-27631, 32716-32728, 34124-34136, 34163-34175, 36832-36844, 38403-38415;

GATA3_03: 10869-10881, 11515-11527, 13845-13857, 17221-17233, 18952-18964, 20050-20062, 40171-40183;

GATA_C: 15848-15860, 18899-18911, 23640-23652, 29072-29084, 30881-30893, 33198-33210, 37472-37484, 38621-38633;

GFI1_01: 35469-35481, 35492-35504;

HFH2_01: 15939-15955, 24636-24652, 25866-25882, 32171-32187, 35372-35388, 39457-35473;

HFH3_01: 13340-13356, 19218-19234, 21328-21344, 21336-21352, 21344-21360, 28062-28078, 32125-32141;

HFH8_01: 14133-14149, 22578-22584;

HNF3B_01: 13150-13166, 16505-16521, 25264-25280, 29443-29459, 37654-37670;

IK2_01: 11547-11559, 17144-17156, 18961-18973, 23883-23895, 27617-27629, 28908-28920, 29241-29253, 30752-30764, 34768-34780;

LYF1_01: 12319-12331, 19191-19203, 37226-37238, 39430-39442;

MAX_01: 22974-22986, 33339-33351;

MZF1_01: 26105-26113, 35187-35195;

NF1_Q6: 12048-12064, 33334-33354;

NFAT_Q6: 13295-13313, 14157-14175, 14311-14329, 14414-14432, 18269-18287. 19326-19344, 20801-20819, 21177-21195, 22537-22555, 23861-23879, 25392-25410, 25879-25897, 27524-27542, 30636-30654, 30718-30736, 31525-31543, 33655-33673, 34726-34744, 34917-34535, 34990-

35008, 35979-35997, 36479-36493, 36577-36595, 37154-37172, 40224-40242, 40365-40383;

NKX25_01: 12041-12055, 12340-12354, 12471-12485, 12742-12756, 12877-12891, 13849-13863, 18995-19009, 21440-21454, 21883-21897, 28426-28440, 30964-30978, 32033-32047, 32265-32279;

NKX25_02: 10998-11012, 12711-12725, 14131-14145, 14726-14740, 16024-16038;

NMYC_01: 18753-18765, 18754-18766, 23076-23088, 30534-30546, 34400-34412;

RORA1_01: 13134-13152, 22966-22984, 24934-24952, 33341-33359, 34760-34778;

S8_01: 11000-11012, 11977-11989, 12048-12060, 12051-12063, 13747-13759, 13923-13935, 13926-13938, 14676-14688, 14679-14691, 16026-16038, 16313-16325, 16316-16328, 17515-17527, 20756-20768, 20759-20771, 23154-23166, 23157-23169, 25198-25210, 25201-25213, 26651-26663, 27508-27520, 27511-27523, 29450-29462, 29478-28490, 29775-29787, 29778-29790, 29813-29825, 29816-29828, 31329-31341, 31677-31689, 31680-31692, 31732-31744, 31735-31747, 36137-36149, 36140-36152, 36812-36824, 36815-36827, 37413-37425, 38679-38691, 39474-39486, 39477-39489;

SOX5_01: 27397-27413, 27572-27588, 28100-28116, 29230-29246, 29439-29455, 30690-30706, 31595-31611, 33871-33887, 34113-34129, 34624-34640, 37668-37684, 38582-38598, 39124-39140, 40410-40426;

SRY_02: 20016-20032, 22410-22426, 27329-27345, 29162-29178, 29499-29515, 30646-30662, 31503-31519, 35928-35944, 37324-37340;

TATA_01: 32722-32738, 32729-32745, 32807-32823, 33825-33841, 34120-34136, 35433-35449, 36593-36609;

TATA_C: 11015-11031, 11817-11833, 13635-13651, 14930-14946;

TCF11_01: 18543-18549, 22574-22580, 31281-31297, 31489-31505, 38754-38770;

USF_01: 23075-23087, 32577-32589;

VMYB_02: 11526-11538, 17384-17396, 18400-18412, 19549-19561, 22188-22200, 40486-40508 and

XFD2_01: 16620-16636. 18153-18169, 22102-22118, 23141-23157.

And a transcription binding site selected from the group consisting of

BINDING SITES

huMDM2, l location in SEQ ID NO:4

AP1_C: 44584-44594, 49069-49079:

AP1_Q2: 42174-42184, 45217-45227, 48422-48422, 50447-50457;

AP1_Q4: 42702-42712, 50806-50816;

AP4_Q6: 42117-42133, 42118-42134, 42244-42260, 45432-45448; 45433-45449, 46609-46625;

BRN2_01: 42310-42328, 44022-44040, 47514-47532, 48900-48918, 48967-48985;

CAAT_01: 44866-44880;

CDPCR3HD_01: 45671-45689, 49219-49237;

CREL_01: 42437-42449, 49797-49809;

FREAC7_01: 47026-47042, 47292-47308, 47658-47674;

GATA1_02: 43482-43494, 48926-48938, 49284-49296;

GATA1_03: 47371-47383;

GATA1_04: 43054-43066, 43162-43162, 43967-43979, 45464-45476, 45916-45928, 47763-47775;

GATA1_05: 49319-49331, 49459-49471;

GATA1_06: 47590-47602;

GATA2_02: 42660-42672, 43475-43487;

GATA2_03: 43714-43726, 50948-50960;

GATA3_02: 49155-49167, 49844-49856;

GATA3_03: 42202-42214, 44810-44822, 48438-48450, 49136-49148, 49337-49349, 49869-49881;

GATA_C: 44011-44023, 45256-45268, 45823-45835, 47915-47927, 49201-49213, 49573-49585;

GFI1_01: 46606-46618, 47063-47075;

HFH3_01: 47030-47046, 47284-47300, 47288-47304;

IK2_01: 45275-45287;

LYF1_01: 44564-44576, 46991-47003, 49567-49579;

MAX_01: 43234-43246, 48726-48738;

MZF1_01: 41772-41780, 42290-42298, 42295-42303, 44507-44515, 45105-45113, 45203-45211, 49948-49956, 50774-50782;

NF1_Q6: 50209-50229;

NFAT_Q6: 42061-42079, 44418-44436, 46399-46417, 47974-47992, 49267-49285, 49964-49982, 50392-50410;

NKX25_01: 42394-42408, 43507-43521, 46115-46129;

RORA1_01: 45073-45091, 48718-48736;

S8_01: 43552-43564, 45214-45226, 47160-47172, 48419-48431, 49295-49307, 50379-50391;

SOX5_01: 43716-43732, 46351-46367, 47156-47172, 47774-47790, 47868-47884, 47974-47990, 48915-48931, 50323-50339;

TATA_01: 45588-45604, 47625-47641, 48026-48042, 48659-48675, 49056-49072, 49079-49095, 49152-49168;

TCF11_01: 49115-49131;

VMYB_02: 42010-42022, 42279-42291, 44651-44663; and

XFD2_01: 42870-42886, 42910-42926.

10. A composition comprising the nucleic acid molecule of claim 7 and a carrier.

15. A kit comprising the nucleic acid molecule of claim 7.

16. The kit according to claim 15, in which the nucleic acid molecule is labeled with a detectable substance.

17. A solid support comprising the nucleic acid molecule of claim 7.

18. The solid support of claim 17 wherein said support is a microarray.

20. The solid support of claim 18, which further comprises a nucleic acid molecule encoding human mouse double minute 2 homolog, complementary sequence thereof or a portion of said nucleic acid molecule containing at least 20 contiguous nucleotides.

22. A method of identifying variants of SEQ ID NO:4, or its complementary sequence, comprising

isolating genomic DNA from a subject and determining the presence or absence of a variant in said genomic DNA using the nucleic acid molecule of claim 7.

23. A method for detecting the presence or absence of SEQ ID NO:4 or its complementary sequence in a sample, said method comprising (a) contacting the sample with the nucleic acid molecule of claim 7 and (b) determining whether the nucleic acid molecule binds to said nucleic acid sequence in the sample.

24. An isolated nucleic acid molecule 20-5000 contiguous nucleotides in length consisting of a reverse or forward strand of a contiguous exon-intron region between nucleotides 41738-9502 of SEQ ID NO:4, or contiguous intron-exon region between nucleotides 41738-9502 of SEQ ID NO:4, wherein a sequence segment comprising 41738-9502 of SEQ ID NO:4 encodes human mouse double minute 2 homolog depicted in SEQ ID NO:2.

25. The isolated nucleic acid molecule of claim 24, wherein said nucleic acid molecule is 20-5000 contiguous nucleotides in length and comprises nucleotides 41739-41738, 40645-40646, 36309-36310, 36384-36385, 32994-32995, 33126-33127, 29564-29565, 29615-29616, 25507-25508, 25287-25288, 25383-25384, 25576-25577, 21006-21007, 21168-21169, 13953-13954, 14109-14110, 13188-13189, 13266-13267, 10664-10665 and/or 9504-9503 of SEQ ID NO:4 or their reverse strands.

30. A microarray comprising a plurality of the nucleic acid molecules of claim 7.

31. The microarray of claim 30 wherein said microarray further comprises a nucleic acid molecule encoding human mouse double minute 2 homolog, complementary sequence thereof or a portion of said nucleic acid molecule containing at least 20 contiguous nucleotides.

32. A method for detecting the presence of a nucleic acid sequence of SEQ ID NO:4 or its complementary sequence in a sample, said method comprising contacting the sample with the nucleic acid molecule of claim 7 and determining whether the nucleic acid molecule binds to

said nucleic acid sequence in the sample.

EVIDENCE APPENDIX

Three Exhibits are included. These include evidence relied on by the Appellant during the Appeal and previously entered by the Examiner. Their location in the

Exhibit 1: Genecard, Anderson et al., 1996, Mammalian Genome 7:780-783 and Bureau, 1995, Genomics 28: 109-112-Submitted with amendment dated August 25, 2005

Exhibit 3: IDS submitted March 1, 2005

Exhibit 4: Liang et al. submitted with amendment dated August 25, 2005

EXHIBIT 1

Netscape Search mdm2 and ge

Return to Results | Remove Frame

GeneCards®

an academic web site of the WEIZMANN INSTITUTE OF SCIENCE in association with XENNEX

Terms of Use | GeneCards Homepage | Search Examples | Comment Form

Notice - Please read carefully prior to linking to any third-party site.

GeneCard for gene MDM2
GC12P067488

Approved UCL/HGNC/HUGO Human Gene Nomenclature database symbol
MDM2 (Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mou

Aliases and Descriptions
(According to GDB, OMIM, HUGO, LocusLink, SWISS-PROT/TrEMBL and/or GeneLoc)

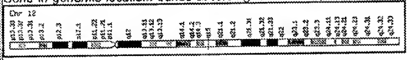
- hdm2 (LL)
- Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse) (GDB, LL)
- mouse double minute 2, human homolog of; p53-binding protein (GDB)
- p53-binding protein MDM2 (LL)
- Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncop

Previous GC identifiers: GC12M068977 GC12P069030 GC12P068918

Chromosomal Location
(According to GeneLoc and/or HUGO, and/or LocusLink (NCBI build 34), Genomic Views According to UCSC and Ensembl)

Chromosome: **12** GeneLoc gene densities
LocusLink cytogenetic band: **12q14.3-q15** *Ensembl cytogenetic band:* **12q15**

Gene in genomic location: bands according to Ensembl, locations according to Gene



GeneLoc location for GC12P067488: (about GC identifiers)

Start: **67,488,238** bp from pter
End: **67,520,481** bp from pter
Size: **32,243** bases
Orientation: **plus** strand

Genomic View:
UCSC Golden Path with GeneCards custom track

Proteins
(According to SWISS-PROT/TrEMBL and/or MIPS, PDB rendering according to OCA)

Swiss-Prot: MDM2_HUMAN

- **Size**: 491 amino acids; 55232 Da
- **Cofactor**: ZINC IS REQUIRED FOR UBIQUITIN LIGASE E3 ACTIVITY.
- **Subunit**: BINDS P53, P73, ARF(P14), RIBOSOMAL PROTEIN L5 AND SPECIFIC WITH RETINOBLASTOMA PROTEIN (RB), E1A-ASSOCIATED PROTEIN P300 AI
- **Subcellular location**: NUCLEAR AND CYTOPLASMIC. EXPRESSED PREDOMI INTERACTION WITH ARF(P14) RESULTS IN THE LOCALIZATION OF BOTH PR NUCLEOLAR LOCALIZATION SIGNALS IN BOTH ARF(P14) AND MDM2 MAY BE NUCLEOLAR LOCALIZATION OF BOTH PROTEINS.
- **Alternative products**: Alternative splicing.
- **Post-translational modifications**: PHOSPHORYLATED IN RESPONSE TO IONI
- **Miscellaneous**: MDM2 RING FINGER MUTATIONS THAT FAILED TO UBIQUITIN P53 FOR DEGRADATION WHEN EXPRESSED IN CELLS.
- **3D structure**: PDB id 1YCR (3D)

MIPS Pedant Viewer: 74112 74106 74107 74108 74109

REFSEQ proteins (6 alternative transcripts): NP_002383.1 NP_006869.1 NP_006870.1 NP_006871.

Protein Domains/ Families
(According to InterPro, ProtoNet, SWISS-PROT/TrEMBL

InterPro Domains and Families:

IPR001876 Znf_RanBP
IPR003121 SWIB
IPR001841 Znf_ring

Mammalian Genome 7, 780–783 (1996).

Mammalian Genome

YAC clones that extend the human Chromosome 12cen-12q15 region contig map

S.E. Andersen, N.J. Lench, A.F. Markham

Molecular Medicine Unit, University of Leeds, Clinical Sciences Building, St James's University Hospital, Leeds LS9 7TF, UK

Received: 4 March 1996 / 5 June 1996

Human Chromosome (Chr) 12q13 is a region of clinical interest in that a variety of disease phenotypes have been localized to the area. A number of relatively detailed genetic maps of the region around human Chr 12q13 are now available. A second-generation YAC contig map of human Chr 12, which extensively covers this area, has also been produced. This has resolved a number of discrepancies in the previous genetic maps. The current YAC contig map contains only five small gaps in the approximately 45 cM region between the polymorphic microsatellite markers *D12S333* and *D12S106*. A number of expressed sequences have also been localized with varying degrees of precision to the region around 12q13.

In an attempt to generate YAC resources with low levels of chimerism and to facilitate disease and EST mapping in this genomic area, we now describe the isolation of 130 YACs from the ICI library (Anand et al. 1990). We have used 36 genetic markers and 11 ESTs from five previous Chr 12 maps (Guyer and Cann, 1992; Weissenbach et al. 1992; Schoenmakers et al. 1994; Kucherlapati et al. 1994; Gyapay et al. 1994). Our data can be integrated with the most recent second-generation YAC contig map of Chr 12 (Krauter et al. 1995). We have mapped the WNT-1 gene (Nusse et al. 1991) at 62–66 cM (based on the 200 cM overall Chr 12 map of Krauter et al. 1995) in close proximity to *D12S339* (in the contig gap between *D12S85* and *D12S361*) and the GPD1 gene. This is an area under-represented in the latest Chr 12 contig map. We have also observed close linkage between GADD153 and GLI at 77 cM in the interval between *D12S312* and *D12S90* (Krauter et al. 1995). The markers *D12S17* and *D12S96* also co-localize at 71 cM in the area between *D12S390* and *D12S398*, which has enabled us to position several first-generation genetic markers with respect to these points. We have isolated YACs for LRP/A2MR, which is known to be physically linked to GLI (Forus and Myklebost 1992). These may be of value in closing a gap at 76 cM in the current map immediately centromeric of GADD153/GLI. The YAC 11GH7 containing *D12S312* may also be useful in this respect as it is the closest known marker proximal to the same contig gap. Similarly, YACs containing MDM2 and *D12S43* may help to close the gap between RAP1B and *D12S80* at 85 cM. As well as YACs containing COL2A1 and LALBA, which are included in the latest contig map, we have isoalted YACs for PMCA, CD63, and PAB1, which are as yet not accurately positioned on the map. *D12S59* or IGF-1 containing YACs described herein are now known to map distal to our region of interest at 113 cM and 118 cM respectively (Krauter et al. 1995). Vectorette YAC end sequencing (Riley et al. 1990) has enabled us to generate other novel STSs and corresponding YACs from the region.

Novel YACs identified with 36 genetic markers from the 12q13 region and from 11 expressed sequences in the region are presented in Table 1. Where known, a corresponding YAC from the second-generation YAC contig map of human Chr 12 is included in Table 1 for ease of cross-reference (Krauter et al. 1995). The 47 loci screened all produced at least one positive YAC and, in most cases, multiple YACs, consistent with a ×3.5-fold genomic representation of this library (Anand et al. 1990). In a number of cases, the same YAC was identified with more than one genetic marker, implying that these markers lie in close physical proximity (the average insert size of the YACs in this library is 350 kb). *D12S339* and the WNT-1 oncogene both gave positive signals with YACs 26BB4, 28AD4, and 32BC2 in the region of the 64 cM contig gap (Fig. 1). *D12S368* and *D12S174* at around 68 cM both identified YAC 30AD11. *D12S17* and *D12S96* both identified the YACs 20BB4 and 20BF12. This places *D12S17* at the 71 cM locus in the middle of the KRT gene family and centromeric of *D12S398*, *D12S359*, *D12S19*, *D12S325*, HOXC5, and *D12S103* respectively. GADD153 and GLI1 at 76 cM identify the YAC 26EG10, confirming the close physical linkage of these two genes. This locus is immediately telomeric of a gap at 76 cM in the current second-generation YAC contig map, and end sequencing of 26EG10 may assist in contig closure. The markers *D12S305* and *D12S104* at 77 cM both identify the YAC 10ED1 concordant with the second-generation map. End sequencing of 10ED1 and rescreening of the YAC library identified 15BE5. This YAC also contains *D12S90*, confirming marker order *D12S90*, *D12S305*, *D12S104*. *D12S83* and *D12S334* both identify the YAC 34CA7, confirming their physical proximity at 77 cM also (Krauter et al. 1995).

YACs for which sequencing of human DNA insert termini has been performed are shown in Table 1. Following confirmation of localization to Chr 12 with the Coriell Mapping Panel, additional YACs were isolated as shown (Table 1). These increase the YAC coverage at *D12S96/17* (71 cM, 20BF12), *D12S90/305/104* (77 cM, 10ED1, 15BE5), and *D12S43* around the 86 cM region (34HH10). A number of YACs proved to be chimeric during analysis; the right-hand end of 40AA12 (a YAC identified using primers from the MDM2 gene) maps to human Chr 1. The MDM2 YAC 6FD11 is also highly chimeric by FISH analysis (data not shown). The MDM2 YAC 40CB5 was not chimeric by FISH analysis and was used to confirm its localization to the 12q14 region (Fig. 2; Heighway et al. 1994). The other YAC termini sequences in Table 1 all map to Chr 12 by PCR analysis of somatic cell hybrids.

In the last four years, six separate Chr 12 maps have been presented which include markers in and around the 12q13 region. The NIH/CEPH collaborative mapping group comprehensive genetic linkage map of the human genome (Guyer and Cann 1992), in conjunction with data of Schoenmakers and associates (1994), provided an approximate marker order cen–COL2A1–ELA1–*D12S29*–*D12S15*–*D12S25*–*D12S14*–*D12S4*–*D12S18*–*D12S16*–*D12S17*–*D12S6*–*D12S28*–*D12S22*–*D12S28*–*D12S19*–*D12S43*–*D12S8*–*D12S64*–*D12S7*–tel, with the latter probably located distally in 12q15. In complementary studies in 1992, Weissenbach and colleagues described another set of markers across the same

Correspondence to: A.F. Markham

Table 1. YACs identified with known genetic markers and ESTs. YACs in bold are detected by more than one marker. CEPH YACs are included for ease of cross reference. YACs identified by rescreening the library with Vectorette-generated end sequences are also included (Riley et al. 1990) as are the novel STSs generated in this process. The NIGMS monochromosomal human/rodent somatic cell-hybrid mapping panels #1 and #2 (NIGMS, Coriell Repository, Camden, N.J., USA) were screened by PCR amplification with newly generated oligonucleotides from YAC insert termini as primers. PCR conditions were as described for the YAC library screening. The panel was screened to confirm localization to human Chr12 and thus to ensure, to a first approximation, that the identified YACs were not chimeric. These sequences have been assigned Genbank accession numbers U51106-U5-1111.

Markers	Tm (°C)	$MgCl_2$ (mM)	Identified ICI YACs	CEPH YACs
D12S17	58.5	1.5	**20BB4; 20BF12**	—
D12S43	55.0	1.8	9GC9; 34HH10	—
D12S58	53.0	2.0	27ID4; 27ID8; 29GE6	—
D12S59	53.0	1.5	18H2; 13GD4	707f6
D12S80	53.0	2.0	5FF10; 18EG1; 19DH6; 32HE12; 34DD7; 37FC5; 38GG8	749a7
D12S83	50.0	1.5	34CA7	799a11
D12S85	53.5	2.0	5IA12; 12HC4; 12IE9; 18CH4; 21GB5, 37EH10, 39DB1	690b1; 767g5
D12S87	52.0	2.0	1HG9; 10IG5; 10IG6; 23BA10; 37AE2, 37FA3	759b2
D12S90	53.0	1.5	4DD8; 9EG8; **15BE5**; 23ID1; 34HA8; 36IC2	770a7
D12S96	53.0	2.4	**20BB4; 20BF12**	934b7
D12S103	53.0	1.5	7DF1	790f7
D12S104	52.0	1.0	**10ED1**; 30AA4; 39EH4	907f2
D12S106	53.0	1.5	33FH8; 35HF4; 39IH2	763a8
D12S118	49.0	1.5	5HB11; 18DA9	—
D12S131	54.5	1.5	pool 8A	—
D12S137	55.0	1.5	40AF6	396e10
D12S174	59.5	1.5	**30AD11**	717g8
D12S305	51.0	1.5	**10ED1**; 20HH9; 38BC9	907f2
D12S312	55.0	1.5	11GH7	790f7
D12S313	51.0	1.5	21AA8	751a4
D12S325	52.0	1.5	8BB12; 8EC4; 24BF5; 39BB1	928a12
D12S326	54.5	1.5	30FE5	806c12
D12S329	56.5	1.5	11GC11; 16GG9, 24FH7	970d8
D12S331	51.5	1.5	9GD6; 11FD5; 40GB12	690h5
D12S333	53.0	1.5	6HG8; 12GG4; 12GH4; 19DD7; 34DD8	952a6
D12S334	49.0	1.5	34CA7; 35ID6	799a11
D12S335	52.5	1.5	pool 15F; 34G; 35G; 39F	755d7
D12S337	55.0	1.5	18AG10; 20EH9; 22EC12; 27BH9	925b12
D12S339	57.5	1.5	5FA12; 9CG7; **26BB4; 28AD4; 32BC2**	—
D12S345	49.0	1.5	19FF10; 29HD4; 29HD11; 33HH12; 37IH1	951b6
D12S347	53.0	1.5	11BD7	951a5
D12S350	57.0	1.5	16BB1; 24CE11; 27EE5; 29CE3; 35CC11; 39HB2; 40CC4	814f12
D12S355	52.5	1.5	39CB8	959a5
D12S361	53.0	1.5	26BG9; 40CD11	951a5
D12S368	53.0	1.5	**30AD11**	717g8
D12S371	55.5	1.5	9CB11; 10IH6; 23CC10; 37HC10	926b3

Genes	Tm (°C)	$MgCl_2$ (mM)	
LALBA	53.5	1.5	38FA5
COL2A1	57.0	3.0	1IA5; 25BF10
GADD153	58.0	1.5	**26EG10**
GLI	53.5	1.5	26EG1, **25EG10**
WNT1	55.0	1.5	**26BB4; 28AD4; 32BC2**
MDM2	56.5	1.5	6FD11; 40AA12; 40CB5; 26DC3
LRP/A2MR	56.0	1.5	13FE9; 15DD6
PAB	52.5	1.5	11GG9
CD63	58.0	1.5	26AB8; 27GE12; 33AG6
PMCA	65.0	3.0	4AF4
IGF1	61.0	2.5	17DF5; 30EC4; 39CF12

Original YAC	Left side YACs	Right YACs
10ED1	30AA4; 39ED3	**15BE5**; 29IE10
20BF12	18EF6; 19IC6; 32DA10; 36EF3; 37DD1; 37IF5	N/D
34HH10	14BC11	N/D

Table 1. *Continued.*

YAC clone	Right end sequences
10ED1 Chrom 12 end	5'-TGC TGC TTC CCC TTT ACA CTT TGG GTC TTA CAT GAT GGG CAC TAT TTA GCG AAC CTA TTA TGG TTG GAG GGG CCT TTA TTG CAG AAT GAA CAC AAT CAG CAG AGG TAG AAA CAT GAG AAA ACA CGT GGC TAT TTG GGA ATA ATA TAT TTA GAA GTG AAA AGC TAT GTT ATA ACT AAG AAG ATT ACC TGT T-3'
40AA12 Chrom 1 end	5'-GGA ATT CAC TTA CTA AAA AAT ACA CAT TTA ATA TAC TAC ATA ATC AAG TTG AGG ATC TAC ATA TTC AGC ATA ATA ATT AAA ATT CTC CAA ACA GTG CCC ACG TTT CAA ATG AAG TTA CAT ATC ACT ATC TTT TAA AGT AAG AAT GCA TCA TTG CAT TGC AGT AGA TCG ACG GAG AGG CAG CGA T-3'

YAC clone	Left end sequences
10ED1 Chrom 12 end	5'-TTC TAT ATC ATT GAG TAA GAT AAA CAG CCA AGA GTC TGT ATT GAT AAA GAA TTG ATC TTT GTA TTT TCC TCA GCT TAC AAA AAT ATT CCA CAT TTT CTC CTA AAA CCT TTA AAC TTT TAG AAA TTA ATT AAA TCT TCC ATC CAC TTC AAA TTT GTG TTC TGT GTA TAA TAT GAA CTA AAG ATT GAG ATT CAT TGC GTC ATA TGT TAT TCA-3'
20BF12 Chrom 12 end	5'-GAA TTC TTC CAG GAT AAA TGC CTG CGA TAG GTG CCA CGA GAA ACC AAT CCT CAC CCT ATT CCC GGA TGC ACT TCA TCA GAG CAA TTA TAG ACT GGT CTG TAT GTA GTC TCA CAG AAG AGC TTT GCT CAG AAG TCA TCT TAG TTA TTA ATT AGA CCA ATA AAA TTC TTG ATG GTC AAA TAG GT-3'
34HH10 Chrom 12 end	5'-AAG CAT TTC CAT GAC CCA CAG AGT CCC CAA GGA ATA GAG TAC AAC TAG GAG ATG AGT CGT GGC ATG GAA GGA AAA GTC AA-3'
40AA12 Chrom 12 end	5'-GTT GCT TTA AGG CCC AAG ACT TTA ATT TAT CAC TAC GGA ATT CAT TTT ATA AGT GAA ATC TGA TCC ACA AAT TTT AAC ACT ATA TTC AGC AAA TGA TAA ACA TAT TTT GCA GC-3'

region from *D12S87* to D12S106, a region at that stage thought to be of around 35 cM. The reported order of markers was cen–D12S87–D12S85–D12S96–D12S103–D12S90–D12S104–D12S83–D12S102–D12S80–D12S92–D12S106–tel. It was not clear at that time how these two different sets of markers overlapped. This was established to some extent in the report of the 2nd International Workshop on Human Chromosome 12 Mapping (Kucherlapati et al. 1994).

The 1993/94 Genethon human genetic linkage map (Gyapay et al. 1994) described 28 genetic markers over an expanded 45 cM region flanked by *D12S87* and *D12S106*. This map was integrated with that of Weissenbach and coworkers (1992). Most recently, a second-generation YAC contig map of human Chr 12 has been produced (Krauter et al. 1995). This represents an almost complete physical map of the region 12cen–12q15 and resolves many of the conflicts concerning accurate ordering of genetic markers. Indeed, the contig map places *D12S87*, previously viewed as a 12q11 marker, on the short arm with *D12S333* even more distal on 12p11. The marker *D12S331* is now positioned at 55 cM just below the centromere on the q arm. However, this latest map still contains five gaps in the region we have been investigating and does not include many first-generation genetic markers, some of which can now be placed on the map.

While the second-generation YAC contig map of the 12q13 region can now be considered as definitive, comparison of the data therein with previous genetic maps and the results described in this paper allows us to draw some useful additional conclusions. Working from centromere to qter, there is an EST (ATP5B) mapping in the 12q11-12 region (Kucherlapati et al. 1994) that is not included in the YAC contig map. This presumably lies between the centromere (54 cM) and the LALBA/*D12S120*/COL2A1/VDR/*D12S85* locus, which is now well resolved on the physical map but was not clearly discriminated on the genetic map. It is therefore possible that ATP5B maps in the first gap in the current YAC contig at around 59 cM. The marker *D12S339*, which was included between *D12S85* and D12S361 (around the 64 cM mark) in the 1993/94 Genethon map, is not included in the 1995 YAC contig map. We have now shown that *D12S339* is closely linked physically to WNT1. It seems probable that these two markers map in the sec-

ond contig gap around 64 cM on proximal 12q. The 1994 Chromosome 12 Workshop map links WNT1 with *D12S131*, *D12S6*, *D12S17*, and *D12S71* (Kucherlapati et al. 1994). Our data contradict these results in that we have demonstrated tight physical linkage between *D12S17* and *D12S96*, the latter now mapping more distally at 71 cM in the middle of the keratin gene cluster. Nevertheless, it would be worth while to screen YAC libraries with *D12S6*, *D12S71* (*D12S131* is reported herein), because it is conceivable that these may also lie in the 64 cM 12q YAC contig gap. The data of Schoenmakers and associates are also valuable in this regard. They localized a series of first-generation Chr 12 markers to the region between COL2A1 and *D12S17*. Their approximate locus order was cen–ELA1-*D12S29*–*D12S15*–*D12S25*–*D12S14*–*D12S4*–*D12S18*–*D12S16*–*D12S17*–tel. They also mapped *D12S6* distal to *D12S17*, which by our data would also place it distal to *D12S96*. Thus, the two genetic maps are somewhat at variance. However, ELA1, *D12S29*, *D12S15*, *D12S25*, *D12S14*, *D12S4*, *D12S18*, and *D12S16* would be potential markers to use to identify YACs in an attempt to fill the second YAC contig gap at 64 cM.

Our physical linkage of *D12S17* and *D12S96*, plus the 1995 localization of *D12S19* at 71 cM (Krauter et al. 1995), would suggest (by comparison of the 1992 NIH/CEPH map with the data of Schoenmakers et al.), that *D12S6*, *D12S28*, and *D12S22* map between *D12S96* and *D12S19*. This can now be easily confirmed because the YAC contig is complete between these points. Again, the 1992 NIH/CEPH map and Schoenmakers and colleagues (1994) have previously placed the markers *D12S43*, *D12S8*, *D12S64*, and *D12S7* distal to *D12S19*. The 1994 workshop report (Kucherlapati et al. 1994) linked *D12S8*, *D12S43*, and LYZ with *D12S80*, which is now positioned on the YAC contig at 87 cM. This is immediately telomeric of a fourth gap in the YAC contig at 85 cM. *D12S8*, *D12S43*, and LYZ are useful markers to screen YAC libraries in an attempt to bridge the 85 cM contig gap between RAP1B and *D12S80*. Furthermore, the 1994 workshop report (Kucherlapati et al. 1994) links the IFNγ and RAP1B loci with *D12S56*. Again, *D12S56* may be useful in filling the gap in this region. Moreover, Bureau and coworkers (1995) have mapped the MDM2 gene in close linkage to IFNγ, but the MDM2 gene is not included on the 1995 contig. It is, therefore, also possible that MDM2 lies in the same contig gap. In support of this, the MDM2 YACs differ from our *D12S313* YAC immediately proximal to IFNγ. This suggests that MDM2 may be distal to IFNγ. While two of the YACs described herein for MDM2 are highly chimeric (6FD11 and 40AA12), the YACs 40CB5 (Fig. 2) and 26DC3 (Heighway et al. 1994) have been localized by FISH analysis to 12q14.3-q15 and may be useful in bridging the contig gap. This cytogenetic localization is in agreement with the reported localization of IFNγ (Bureau et al. 1995) and with the mapping of the slightly distal markers *D12S213* and *D12S115* to 12q14-q15 by Fejzo and associates (1995). It is interesting that the previously noted co-amplification of the GLI, CDK4 and MDM2 genes in human sarcomas (Khatib et al. 1993) is entirely consistent with the genetic map of this region. The YACs for *D12S350* and *D12S326* may prove useful in bridging a fifth gap, which they flank at 91 cM in the current contig map.

D12S64 may be some way distal to the *D12S8*/*D12S43*/*D12S80* locus at 12q14-q15, and in fact *D12S7* maps at 103 cM, considerably distal to the *D12S106* marker at 98 cM which we have taken as the boundary of our area of investigation. Similarly, *D12S58*, for which three YACs are described herein, now maps at 113 cM, beyond the 12q15 region. At the centromeric end of the region, *D12S59* and *D12S345*, both map adjacent to the centromere on the short arm of Chr 12 (Kucherlapati et al. 1994; Krauter et al. 1995). Our *D12S87* YACs also map on the p arm, as do our most distal YACs for *D12S333*(12p11).

Of the ESTs that have been used in this study to generate YAC clones, LALBA, COL2A1, GADD153, and GLI are already positioned on the 1995 YAC contig map at around 60, 61, 77, and 77 cM respectively. The IGF1 gene, for which we have identified a number of YACs, maps distal to *D12S106* at around 118 cM. These YACs are therefore included in Table 1 only for archival purposes. Of the ESTs for which we have isolated YACs that are not yet positioned on the 1995 YAC contig map, the WNT1 clones promise to be of value in that they map with *D12S339* in a current contig gap at 64 cM. Similarly, our MDM2 YAC clones may map in the current YAC contig gap at 85 cM. At present, the precise positions of PMCA, CD63, and PAB1 on the physical map remain to be confirmed. Finally, we have isolated YACs for the LRP/A2MR gene (Paulien et al. 1992), which has been physically mapped 300 kb from GLI (Forus and Myklebost 1992). These may be useful in closing a third YAC contig gap in this region between *D12S312* and GADD153 at 76 cM, as indeed may the *D12S312* and GADD153/GLI YACs described herein. The availability of a wide range of YAC reagents from the 12cen-12q15 region will without question facilitate gene cloning exercises in this area of the genome and should allow us to resolve some of the outstanding

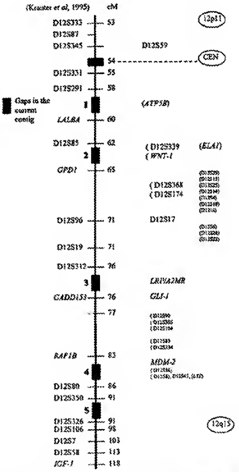


Fig. 1. Schematic 12p11-12q15 genetic map highlighting gaps in the current YAC contig and the positioning of additional markers for some of which new YACs are described herein. The ICI YAC library (Anand et al. 1990) was screened by the polymerase chain reaction (PCR) with available genetic markers from the region. Standard PCRs contained 100 ng pooled YAC DNA solution, 30 pmol of each primer, 1.0 × Taq buffer (Promega), 0.2 mM dNTPs (Promega), 2u*Taq* DNA polymerase (Promega), and 1.5 mM$MgCl_2$ except in specific cases shown in Table 1 where higher $MgCl_2$ concentrations were necessary. All PCRs were carried out with a Techne PHC-3 thermal cycler. The denaturing step was 5 min/95°C for purified DNA; 10 min/95°C for whole yeast cells; followed by 38 cycles of 30 s/95°C, 30 s/primer Tm and 1 min/72°C, with a final elongation step of 10 min/72°C.

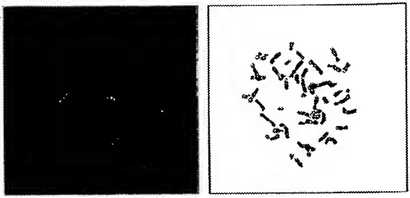


Fig. 2. Fluorescent in situ hybridization analysis of the MDM2 YAC 40CB5 confirming localization of this gene to 12q14-15.

questions concerning the frequent translocations encountered in the region in a variety of malignancies.

Acknowledgments. Research in the authors' laboratories is supported by the Medical Research Council, the Wellcome Trust, the Yorkshire Cancer Research Campaign and the West Riding Medical Research Trust. We thank the Candlelighters for additional provision of specific research support for this project, J.P. Leck for FISH analysis, C.A. Higgins for manuscript preparation and Zeneca plc for library access.

References

Anand, R., Riley, J.H., Butler, R., Smith, J.C., Markham, A.F. (1990). A 3.5 genome equivalent multi-access YAC library: construction, characterisation, screening and storage. Nucleic Acids Res., 18, 1951–1956.

Bureau, J.F., Bihl, F., Brahic, M., Le Paslier, D. (1995). The gene coding for interferon-γ is linked to the D12S335 and D12S313 microsatellites and to the MDM2 gene. Genomics 28, 109

Fejzo, M.S., Yoon, S.J., Montgomery, K.T., Rein, M.S., Weremowicz, S., Krauter, K.S., Dorman, T.E., Fletcher, J.A., Mao, J., Moir, D.T., Kucherlapati, R.S., Morton, C.C. (1995). Identification of a YAC spanning the translocation breakpoints in uterine leiomyomata, pulmonary chondroid hamartoma, and Lipoma: physical mapping of the 12q14-q15 breakpoint region in uterine leiomyomata. Genomics, 26, 265–271.

Forus, A., Myklebost, O. (1992). A physical map of a 1.3-Mb region on the long arm of chromosome 12, spanning the GLI and LRP loci. Genomics, 14, 117–120.

Guyer, M., Cann, H. (1992). A comprehensive genetic linkage map of the human genome: NIH/CEPH Collaborative Mapping Group. Science (reprint) pp 3–66.

Gyapay, G., Morisette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernadi, G., Lathrop, M., Weissenbach, J. (1994). The 1993–94 Genethon human genetic linkage map. Nature Genet. 7, 246–339.

Heighway, J., Mitchell, E.L.D., Jones, D., White, G.R.M., Santibanez Koref, M.F. (1994). A transcribed polymorphism and sub-localisation of MDM2. Hum. Genet., 93, 611–612.

Khatib, Z.A., Matsushime, H., Valentine, M., Shapiro, D.N., Sherr, C.J., Look, A.T. (1993). Coamplification of the *CDK4* gene with *MDM2* and *GLI* in human sarcomas. Cancer Res., 53, 5535–5541.

Krauter, K., Montgomery, K., Yoon, S.J., LeBlanc-Straceski, J., Renault, B., Marondel, I., Herdman, V., Cupelli, L., Banks, A., Lieman, J., Menninger, J., Bray-Ward, P., Nadkarni, P., Weissenbach, J., LePaslier, D., Rigault, P., Chumakov, I., Cohen, D., Miller, P., Ward, D., Kucherlapati, R. (1995). A second-generation YAC contig map of human chromosome 12. Nature, 377, S321–S333.

Kucherlapati, R., Craig, I., Marynen, P. (1994). Report of the second international workshop on human chromosome 12 mapping 1994. Cytogenet. Cell Genet. 67, 245–276.

Nusse, R., Brown, A., Papkoff, J., Scambler, P., Shackleford, G., McMahon, A., Moon, R., Varmus, H. (1991). A new nomenclature for int-1 and related genes: the Wnt gene family. Cell, 64, 231–232.

Paulien, S., Sandberg, A.A., Herz, J., Gemmill, R.M. (1992). Putative apolipoprotein receptor gene (LRP, A2MR) is not rearranged in either myxoid liposarcoma or lipomas with translocations in 12q13-14. Cancer Genet. Cytogenet., 60, 125–130.

Riley, J.H., Butler, R., Ogilvie, D., Finniear, R., Jenner, D.E., Powell, S., Anand, R., Smith, J.C., Markham, A.F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res., 18, 2887–2890.

Schoenmakers, E.F.P.M., Kools, P.F.J., Mols, R., Kazmierczak, B., Bartnitzke, S., Bullerdiek, J., Dal Cin, P., De Jong, P.J., Van Den Berghe, H., Van De Ven, W.J.M. (1994). Physical mapping of chromosome 12q breakpoints in lipoma, pleomorphic salivary gland adenoma, uterine leiomyoma and myxoid liposarcoma. Genomics, 20, 210–222.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M. (1992). A second-generation linkage map of the human genome. Nature 359, 794–801.

SHORT COMMUNICATION

The Gene Coding for Interferon-γ Is Linked to the D12S335 and D12S313 Microsatellites and to the MDM2 Gene

J. F. BUREAU,* F. BIHL,* M. BRAHIC,*,[1] AND D. LE PASLIER†

*Unité des Virus Lents, Institut Pasteur, URA CNRS 1157, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France; and
†Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France

Received March 13, 1995; accepted April 27, 1995

Interferon-γ is a cytokine with multiple effects. It interferes with the replication of several viruses and plays a key role in the regulation of immune responses. Therefore, the gene coding for interferon-γ could be implicated in the susceptibility of humans to several diseases. We have localized this gene close to the D12S335 and D12S313 microsatellites on both the physical and the genetic maps of the human genome. We also physically mapped this gene close to the MDM2 locus on chromosome band 12q15. Finally, we describe the organization of the *Ifg*, *Myf-6*, *Mdm1*, and *Mdm2* loci on mouse chromosome 10, in a region syntenic to human chromosome band 12q15. © 1995 Academic Press, Inc.

Interferon-γ (IFNG) was discovered because of its antiviral property (12), although it is now studied chiefly because of its central role in the regulation of immune responses. This cytokine is secreted by $CD4^+$ T cells committed to the Th1 pathway, by $CD8^+$ T cells, and by activated macrophages. Because of these multiple functions, the gene coding for IFNG is a good candidate gene for control of susceptibility to various infectious as well as immune-mediated diseases of humans and other mammals. The persistent infection of the mouse central nervous system by Theiler's virus is a case in point. The *Ifg* gene is a good candidate for control of the persistence of the infection for two reasons: (i) Persistence is controlled by a gene that was mapped to the telomeric region of chromosome 10, close to the *Ifg* locus (5). (ii) Resistant 129Sv mice whose gene coding for the IFNG receptor has been inactivated become susceptible (8).

A precise localization of the human IFNG gene was not available until the present work. This gene had been localized to band 12q24 using FISH and by screening a panel of somatic hybrid cell lines (14, 19). Recently, Ruiz-Linares (17) described a microsatellite in the first intron of this gene, and Awata *et al.* (3) reported a difference in the allelic distribution of this marker between a group of patients with insulin-dependent diabetes mellitus and a control group. For future studies using the candidate gene approach, it would be extremely useful to locate precisely the IFNG gene in particular with respect to polymorphic microsatellites markers.

With this goal in mind, we screened the YAC library from CEPH (1) by PCR using two primers (forward 5'-GCTGTTATAATTATAGCTTGT-3' and reverse 5'-AGGGTATTATTATACGAGCT-3') derived from those described by Ruiz-Linares (17) for the IFNG microsatellite. PCR was performed using a kit from Amersham. After denaturation at 94°C for 2 min, 200 ng of DNA in 25 μl was submitted to 40 cycles of amplification (94°C, 40 s; 50°C, 40 s; 72°C, 15 s). Using this procedure, we isolated clone 825G7 of the YAC library. We found that clone 825G7 belonged to a contig of 25 other YAC clones, according to inter-*Alu* PCR patterns described in the CEPH–Généthon WWW server (6). All clones in this contig contained either both microsatellites D12S335 and D12S313 or one of the two. Each clone was tested for the presence of the IFNG microsatellite. Eleven clones contained the IFNG microsatellite. Ten also contained both the D12S335 and D12S313 microsatellites (YAC clones 745A10, 743E2, 751A4, 809H4, 823D1, 870H3, 924E4, 926A6, and 983H8), and one, YAC clone 763F1, contained only the D12S313 microsatellite. Therefore, these results demonstrate that the IFNG gene is physically linked to the D12S335 and D12S313 microsatellites.

To confirm the position of the IFNG locus on the genetic map, we analyzed the segregation of alleles of five microsatellite markers in the eight families that have been used to construct the Généthon map (10). Besides markers D12S335 and D12S313, we used two more microsatellites located on either side of the D12S335–D12S313 region. One was taken from a group of six cosegregating markers (D12S104, D12S305, D12S355, D12S334, D12S83, and D12S329) centromeric to the D12S335 marker and the other from a group of three cosegregating markers (D12S344, D12S80, and D12S92) telomeric to the D12S313 marker. For each family, a pair of markers was chosen according to its polymorphism within the family. The parental meiosis were not informative for

[1] To whom correspondence should be addressed. Fax: 33 (1) 40 61 31 67. E-mail: mbrahic@pasteur.fr.

GENOMICS 28, 109–112 (1995)
0888-7543/95 $12.00

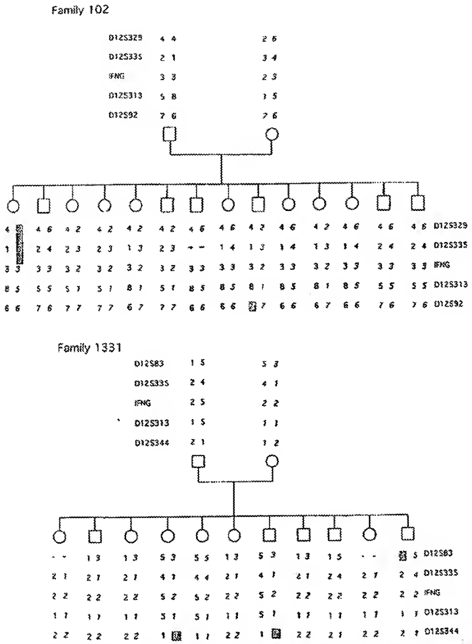


FIG. 1. Segregation of the alleles of IFNG and four microsatellite markers in two families. Maternal alleles are shown italicized. The genotypes of the microsatellites were obtained from the Généthon database. Dashes correspond to cases for which genotypes were not available. Stippled areas indicate a recombination event, and bold face identifies alleles from each parent. The alleles of IFNG are numbered according to the size of the amplified DNA, from the smaller to the larger fragment.

two families. The pedigrees of the other six were analyzed for the markers described above: IFNG, D12S335, D12S313, and the two flanking microsatellites. The results are presented in Fig. 1 for two families and summarized in Table 1. No recombination was observed between the D12S313 and the IFNG markers among 137 informative meioses. Only one recombination was observed between the D12S335 and the IFNG markers among 135 informative meioses. Analysis revealed that the flanking markers were indeed less linked to the IFNG locus than the D12S335 and D12S313 markers.

We next showed, by physical and genetic mapping, that the IFNG locus is linked to the D12S313 and D12S335 loci. Indeed, YAC clone 926A6 was recognized by these three markers. This YAC clone is chimeric and contains regions of both chromosome 12 (band

TABLE 1

Genetic Localization of IFNG

	No. crossing over with IFNG	No. of informative meioses
D12S329/D12S83	5 (3.9 cM)	129
D12S335	1 (0.8 cM)	135
D12S313	0 (0.0 cM)	137
D12S344/D12S92	6 (4.4 cM)	137

12q15) and chromosome 5, as shown by FISH (6, 9). Therefore, our results localize the IFNG locus to band 12q15, which is different from the previously published localization (band 12q24) (19). The region of human chromosome 12 to which we localize the IFNG gene is syntenic to a region of the mouse genome where we and others have previously localized the mouse *Ifg* gene (2, 5, 18). In our work, the positions of loci *Mdm1*, *Ifg*, and *Myf6* and of three microsatellites D10Mit10, D10Mit14, and D10Mit164 from the Whitehead Institute were determined using the progeny of a F1 (B10.S × SJL/J) × B10.S backcross (Fig. 2) (5). This backcross was typed for more than 100 loci that were distributed among all of the autosomal chromosomes. The results showed that the *Mdm1*, *Ifg*, and *Myf6* loci were linked to each other in the telomeric region of chromosome 10. Other authors have also reported that the *Mdm1*, *Mdm2*, *Mdm3*, and *Ifg* loci are closely linked (2, 18). Interestingly, the MDM2 locus has been localized by FISH to human chromosome 12q14.3–q15 (11). Thus, the analysis of the syntenic region of the mouse genome agrees with our localization of the IFNG locus to human chromosome 12q15.

We confirmed the physical linkage of the MDM2 and IFNG loci by screening, with the MDM2 PCR (11), the contig of YAC clones recognized by at least one of the D12S335, D12S313, and IFNG microsatellites. The MDM2 marker recognized three YAC clones (751A4, 870H3, and 983H8) that were also recognized by the D12S335, D12S313, and IFNG microsatellites. Thus, the MDM2 and IFNG loci are physically linked. The MDM2 gene, which codes for a p53-like protein, is located in a region associated with several cancers (15, 16). New polymorphic markers for this region are of particular interest since the only one available so far is a *Nla*IV polymorphic restriction site (11).

The *Myf6* locus is located in the region of mouse chromosome 10 syntenic to human chromosome band 12q15 (4). Therefore, we decided to localize the MYF-6 locus physically. We screened the CEPH YAC library using MYF-6-specific primers (forward 5′-AGACCTTCTCC-ACGCAGCAG-3′ and reverse 5′-GCGAAATCTGTTG-TGCAGCT-3′) under PCR conditions identical to those described above for the MDM2 marker. Two clones, 921C6 and 982A6, were isolated. We found that both belonged to a contig of nine other YAC clones, according to inter-*Alu* PCR patterns described in the CEPH–Généthon WWW server (6). Three of them, clones 940B8, 937D9, and 949H4, contained the MYF-6 marker. According to the CEPH–Généthon server, clones 940B8 and 937D9 contained the D12S106 microsatellite, a marker 15 cM from the D12S313 microsatellite, toward the telomere (10). Figure 3 shows the position of these various markers on human chromosome 12q15 and mouse chromosome 10. The organization of the syntenic regions is very similar. IFNG/D12S313 and D12S106 are 15 cM apart, whereas Ifg/Mdm1 and Myf6 are 5 cM apart. However, since 1 cM is, on average, equivalent to 1.7 Mb in the mouse and 1.0 Mb in human, the physical distances between the markers in mouse and in human could be of the same order.

In conclusion, we have shown that the IFNG and the MDM2 genes are close to each other and to the D12S335 and D12S313 microsatellites. These markers are most likely localized to chromosome bands 12q14.3–q15, in a region syntenic to the telomeric part

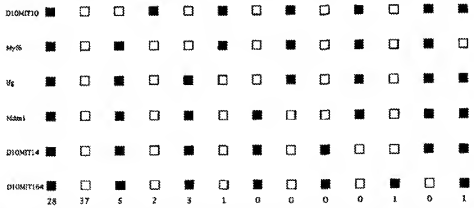


FIG. 2. Segregation of the alleles of *Ifg*, *Mdm1*, *Myf6*, and some markers published by the Whitehead Institute (7) in 78 F1 (SJL/J × B10.S) × B10.S mice. The *Ifg* and *Mdm1* markers were described in Bureau *et al.* (5). A microsatellite was found in the second intron of *Myf6* and used to design two PCR primers (forward 5′-GAAAGGGCACTGGGCTGTAC-3′ and reverse 5′-CGCCGGATTTGGCTGTTGCT-3′) (13). Black squares represent homozygous mice; white squares represent heterozygous mice. Numbers under each column represent the number of mice with each genotype.

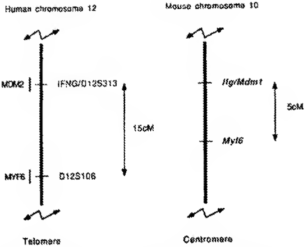


FIG. 3. Genetic maps of the syntenic regions of human chromosome 12q15 and mouse chromosome 10. The MDM2 and MYF6 genes were localized by physical mapping only. The other human genes and markers were localized by both physical and genetic mapping.

of mouse chromosome 10. This information will be extremely useful in human genetic studies in which the IFNG and MDM2 genes will be candidate genes.

ACKNOWLEDGMENTS

We thank C. Dib and J. Weissenbach for their help; C. Petit for helpful discussion; and M. Gau for preparing the manuscript. This work was supported by grants from the Institut Pasteur Fondation, the ARSEP, the CNRS, the GREG, and the NMSS.

REFERENCES

1. Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87:** 4256–4260.
2. Ashar, H. R., Benson, K. F., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., and Chada, K. K. (1994). Ifg, Gli, Mdm1, Mdm2, and Mdm3: Candidate genes for the mouse pg locus. *Mamm. Genome* **5:** 608–611.
3. Awata, T., Matsumoto, C., Urakami, T., Hagura, R., Amemiya, S., and Kanazawa, Y. (1994). Association of polymorphism in the interferon gamma gene with IDDM. *Diabetologia* **37:** 1159–1162.
4. Braun, T., Bober, E., Winter, B., Rosenthal, N., and Arnold, H. H. (1990). Myf-6, a new member of the human gene family of myogenic determination factors: Evidence for a gene cluster on chromosome 12. *EMBO J.* **9:** 821–831.
5. Bureau, J. F., Montagutelli, X., Bihl, F., Lefebvre, S., Guénet, J. L., and Brahic, M. (1993). Mapping loci influencing the persistence of Theiler's virus in the murine central nervous system. *Nature Genet.* **5:** 87–91.
6. Cohen, D., Chumakov, I., and Weissenbach, J. (1993). A first-generation physical map of the human genome. *Nature* **366:** 698–701.
7. Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C., and Lander, E. S. (1994). A genetic map of the mouse with 4006 simple sequence length polymorphisms. *Nature Genet.* **7:** 220–225.
8. Fiette, L., Aubert, C., Müller, U., Huang, S., Aguet, M., and Brahic, M. (1995). Theiler's virus infection of 129Sv mice that lack the interferon alpha/beta or interferon gamma receptors. *J. Exp. Med.* **181.**
9. Francke, U. (1994). Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet. Cell Genet.* **65:** 206–219.
10. Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. (1994). The 1993–94 Généthon human genetic linkage map. *Nature Genet.* **7:** 246–339.
11. Haighway, J., Mitchell, E. L. D., Jones, D., White, G. R. M., and Santibanez Koref, M. F. (1994). A transcribed polymorphism and sub-localisation of MDM2. *Hum. Genet.* **93:** 611–612.
12. Isaacs, A., Lindenmann, J., and Valentine, R. C. (1957). Virus interference. II. Some properties of interferon. *Proc. R. Soc. London B* **147:** 268–273.
13. Miner, J. H., and Wold, B. (1990). Herculin, a fourth member of the myoD family of muscle regulatory genes. *Proc. Natl. Acad. Sci. USA* **87:** 1089–1093.
14. Naylor, S. L., Sakaguchi, A. Y., Shows, T. B., Law, M. L., Goeddel, D. V., and Gray, P. W. (1983). Human immune interferon gene is located on chromosome 12. *J. Exp. Med.* **57:** 1020–1027.
15. Oliner, J. D., Kinzler, K. W., Meltzer, P. S., George, D. L., and Vogelstein, B. (1992). Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature* **358:** 80–83.
16. Reifenberger, G., Lui, L., Ichimura, K., Schmidt, E. E., and Collins, V. P. (1993). Amplification and overexpression of the MDM2 gene in a subset of human malignant gliomas without p53 mutations. *Cancer Res.* **533:** 2736–2739.
17. Ruiz-Linares, A. (1993). Dinucleotide repeat polymorphism in the interferon-gamma (IFNG) gene. *Hum. Mol. Genet.* **2:** 1508.
18. Taylor, B. A., Rowe, L., and Grieco, D. (1992). Close linkage of Mdm-1, a gene amplified and overexpressed in a transformed 3T3 cell line, with gamma interferon (Ifg) on Chromosome 10 of the mouse. *Mamm. Genome* **3:** 700–704.
19. Trent, J. M., Olson, S., and Lawn, R. M. (1982). Chromosomal localization of human leukocyte, fibroblast, and immune interferon genes by means of *in situ* hybridization. *Proc. Natl. Acad. Sci. USA* **79:** 7809–7813.

EXHIBIT 3

JR-10,003-US PATENT

OIPE JC106 MAR 0 3 2005 PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ryan

Serial No.: 10/608,463 Group Art Unit: 1652

Filed: June 27, 2003 Examiner: E. Slobodyansky

FOR: ISOLATED GENOMIC POLYNUCLEOTIDE FRAGMENTS FROM CHROMOSOME 12 THAT ENCODE HUMAN CARBOXYPEPTIDASE M AND THE HUMAN MOUSE DOUBLE MINUTE 2 HOMOLOG

Confirmation No.: 6428

INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with 37 C.F.R. 1.56, 1.97 and 1.98, Applicants submit herewith references which they believe may be material to the patentability of this application and with respect to which there may be a duty to disclose in accordance with 37 C.F.R. 1.56.

While the references may be "material" under 37 C.F.R. 1.56, it is not intended to constitute an admission that the references are "prior art" unless specifically designated as such.

The filing of this Information Disclosure Statement shall not be construed as a representation that no other material references than those listed exist or that a search has been conducted.

The references are listed in PTO form 1449 which is in accordance with the requirements of M.P.E.P. 609. A copy of the references is also enclosed.

The references are as follows:

U.S. Patent Documents

03/04/2005 EAREGAY1 00000022 10608463
01 FC:1806 180.00 OP

Foreign Patent Documents

Other Documents

1. RIES et al., 2000, Cell 103:321-330
2. REHLI et al., 1995, J. Biol. Chem. 270:15644-9
3. OLINER et al., 1992, Nature 358: 80-3
4. TAN et al., 1989, J. Biol. Chem. 264: 13165-70

It is respectfully requested that these references be considered by the Patent and Trademark Office in its examination of the above-identified application and be made of record therein. The Examiner is also invited to contact the Undersigned if there are any questions concerning this paper or the attached references.

The Information Disclosure Statement submitted herewith is being filed

[] before the mailing date of a first Office Action on the merits

[x] after the mailing date of a first Office Action on the merits. Please charge the fee of $180 to credit card. Form PTO-2038 is attached.

[] An International Search Report is enclosed. References cited in the International Search report are asterisked.

Respectfully submitted,

Date: 3/1/05

Cheryl H. Agris

Cheryl H. Agris, Reg. No. 34,086
P.O. Box 806
Pelham, N.Y. 10803
(914) 712-0093

PTO/SB/08A (04-03)
Approved for use through 04/30/2003. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 1 of 2

Complete if Known	
Application Number	10/608,463
Filing Date	June 27, 2003
First Named Inventor	Ryan
Art Unit	1652
Examiner Name	E. Slobodyansky
Attorney Docket Number	JR10003

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			
		US-			

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3] Number[4] Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached.

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: **Commissioner for Patents, Washington, DC 20231.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

OIPE JC106
MAR 0 3 2005
PATENT & TRADEMARK OFFICE

PTO/SB/08B (04-03)
Approved for use through 04/30/2003. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO	Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	**Application Number**	10/608,463
	Filing Date	June 27, 2003
	First Named Inventor	Ryan
(Use as many sheets as necessary)	Art Unit	1652
	Examiner Name	E. Slobodyansky
Sheet 2 of 2	Attorney Docket Number	JR10003

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
		RIES et al., 2000, Cell 103: 321-330.	
		REHLI et al., 1995, J. Biol. Chem. 270: 15644-15649.	
		OLINER et al., 1992, Nature 358:80-83.	
		TAN et al., 1989, J. Biol. Chem. 264: 13165-13170.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached.

EXHIBIT 4

Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 338 (2004) 217–223

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Genomic organisation of the human *MDM2* oncogene and relationship to its alternatively spliced mRNAs☆

Huiling Liang[a,1], Helen Atkins[a], Rana Abdel-Fattah[a], Stephen N. Jones[b], John Lunec[a,*]

[a] *Cancer Research Unit, Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK*
[b] *University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA*

Received 26 January 2004; received in revised form 26 April 2004; accepted 17 May 2004
Available online 21 July 2004
Received by A.J. van Wijnen

Abstract

The *MDM2* proto-oncogene, which encodes a protein that binds to the p53 tumour suppressor, has been found amplified and overexpressed in a range of human tumours. Although the human *MDM2* cDNA sequence has been reported, the genomic organisation of the human gene has not been documented. We have previously reported the detection of five alternative internally deleted *MDM2* transcripts in human tumours and suggested these may represent alternatively spliced forms. Here we demonstrate two novel *MDM2* transcripts with internal deletions, using RT-PCR followed by sequencing. To definitively ascribe these variant transcript forms to alternative splicing, and to explore associated mechanisms, we have determined the intron–exon organisation of the human genomic sequence. The human *MDM2* gene spans approximately 33 kb and is divided into 12 exons. Exon sizes range from 50 to ≥ 1161 bp and intron sizes vary from 121 to ~ 7000 bp. The positions of intron–exon boundaries are compared with the deletion junctions of the multiple-sized transcripts and discussed in relation to alternative splicing mechanism.

Keywords: p53; genomic mapping; long range PCR

1. Introduction

The *mdm2* proto-oncogene was initially identified as an amplified gene from a mouse double minute chromosome present in a spontaneously transformed Balb/C 3T3 cell line, 3T3DM (Haines et al., 1994; Sigalas et al., 1996; Steinman et al., 2004). The causal role of this gene in tumorigenesis was originally established by transfection studies using genomic DNA sequences. In these studies, experimental overexpression of *mdm2* resulted in the immortalisation of primary rat embryo fibroblasts and induced a fully transformed phenotype in the cells when cotransfected with an activated *ras* gene (Finlay, 1993). The human homologue of the *MDM2* gene has been found to be amplified in over 30% of human sarcomas (Oliner et al., 1992; Leach et al., 1993), which consequently results in high levels of the *MDM2* gene product. In addition, *MDM2* overexpression can also occur through enhanced transcription and translation (Bueso-Ramos et al., 1993; Landers et al., 1997; Momand et al., 1998).

The human *MDM2* gene has been localised to chromosome 12q13–14. Although the human *MDM2* cDNA sequence has been previously reported (Oliner et al., 1992), little is known about its genomic organisation. The MDM2 protein is composed of 491 amino acids and contains a p53 binding domain (codons 19–102), a putative nuclear localisation signal (codons 181–185), an acidic domain (codons 223–274), a central zinc-finger motif (codons 305–332) and a ring-finger motif towards the C-terminal end of the protein (codons 438–478) (Boddy et al., 1994).

Abbreviations: *MDM2*, mouse double minute 2 gene; RT-PCR, reverse transcription polymerase chain reaction; cDNA, complementary DNA; kb, kilobase; bp, base pair; TE, tris(hydroxymethyl)aminomethane ethylene diamine tetra acetate.

☆ Database accession numbers: AF144014-AF144033 and AF201370-AF201371.

* Corresponding author. Northern Institute for Cancer Research Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK. Tel.: +44-191-246-4420, fax. +44-191-246-4301.

E-mail address: john.lunec@newcastle.ac.uk (J. Lunec).

[1] Present address: University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.05.015

MDM2 appears to be a pluripotential oncoprotein, exerting its transforming properties through several alternative mechanisms, of which the most extensively studied has been the negative regulation of p53 function. MDM2 blocks p53 transcriptional function by binding to p53 (Momand et al., 1992; Oliner et al., 1993). The binding of MDM2 to p53 also results in the rapid degradation of p53 (Haupt et al., 1997; Kubbutat et al., 1997). In addition, MDM2 has been reported to have p53-independent tumorigenic properties. This includes the ability to interact with and inactivate the pRb tumour suppressor protein (Xiao et al., 1995) and to bind to and activate the E2F1 transcription factor (Martin et al., 1995). Furthermore, two independent transgenic studies have shown *MDM2* to have tumourigenic properties in p53 null mice (Lundgren et al., 1997; Jones et al., 1998).

One of the distinctive properties of *MDM2* is the possession of an extremely complex expression pattern. Its multiple-sized transcripts and proteins have been found in tumour samples and cell lines by a number of groups (Haines et al., 1994; Sigalas et al., 1996; Bartel et al., 2002). In our previous studies, five alternatively sized transcripts of the human *MDM2* were found in human ovarian tumour, bladder tumour and leukaemic cell samples (Sigalas et al., 1996). The expression of the alternatively sized forms was found to be more frequent in tumours of advanced stage and high histological grade, and they also retained their ability to transform NIH3T3 cells. Here, we present data demonstrating two further *MDM2* transcript forms with internal sequence deletions in human tumour tissue. We hypothesised that these transcripts are generated by alternative splicing. To test this hypothesis and to explore the associated mechanisms, we have investigated the genomic structure and organisation of the human *MDM2* gene. This gene is ~33 kb in length and comprises at least 12 exons. The sizes of exons vary from 50 to ≥1161 bp, and introns range in size from 121 to ~7000 bp. The position of intron–exon boundaries is compared with the sequences of the *MDM2* variant transcripts and discussed in relation to alternative splicing mechanisms.

2. Materials and methods

2.1. Nested RT-PCR

Total RNA was extracted from human bladder tumour and normal bladder tissues. Nested RT-PCR was carried out as previously described (Sigalas et al., 1996).

2.2. Genomic DNA extraction

Genomic DNA was prepared from frozen normal human placental tissue by digestion with proteinase K and phenol–chloroform extraction. The DNA was precipitated with a half volume of 7.5M ammonium acetate and one volume of isopropanol, washed in 70% ethanol and resuspended in 1 × TE buffer (10 mM Tris, 0.1 mM EDTA pH 7.5).

2.3. Long-range PCR

Primers (Table 1) were designed from the published *MDM2* cDNA sequence (Oliner et al., 1992). Each primer pair was designed to cross the deletion junctions of multiple-sized transcripts (Sigalas et al., 1996) or according to the predicted exon/intron boundaries by referring to the mouse *mdm2* gene structure (Jones et al., 1996; Montes de Oca luna et al., 1996). A long-range PCR protocol was carried out, with the above human genomic DNA as a template, using an XL PCR Kit (Perkin Elmer, Part No. N808-192). For comparison, PCR was also carried out on normal human placental cDNA. The reaction contained 1 × reaction buffer, 0.8 mM dNTP, 1.1 mM $Mg(OAc)_2$ and 4 units of *rTth* DNA polymerase, 40 pmol of each primer and 100 ng of genomic DNA or 20 μl cDNA in a total volume of 100 μl. The long-range PCR was performed using a thermal cycler (Perkin-Elmer Model 480) as follows: 94 °C for 2 min; cycles 1–16 at 94 °C for 30 s, 58–62 °C for 10 min; cycles 17–28 at 94 °C for 30 s, 58–62 °C for 10 min and 15 s of increment

Table 1
The sequences of primer pairs for amplification of genomic nucleotides of the human *MDM2* gene

Primer name	Forward primer (5′→3′)	cDNA position (nt)[a]	Reverse primer (5′→3′)	Position at cDNA (nt)[a]	Annealing temperature (°C)
Pri1	cctgtgggccctgtgtgtc	30–49	tgttccgaagctggaatctgtg	382–361	60
Pri2	gtgcaataccaacatgtctg	314–333	caacagactttaataacttcaaaagc	435–410	60
Pri3	gcttttgaagttattaaagtctgttg	410–435	tacaatatgttgttgcttctcatc	536–513	57
Pri4	ttttatcttggccagtatattatg	474–497	atcctgctgattgactactacc	654–632	57
Pri5	catgatctacaggaacttggtag	614–636	ttgatcactcccaccttcaagg	716–695	60
Pri6	actcaggtacatctgtgagtgag	661–683	tgtctcactaattgctctccttc	815–793	60
Pri7	ttgtacaagagcttcaggaagag	724–746	atggcgtccctgtagattcac	969–949	60
Pri8	tgaaagcctggctctgtgtg	893–912	caaattctacactaaactgatctg	1059–1036	57
Pri9	tcttgatgctggtgtaagtgaac	980–1002	agctaaggaaatttcaggatcttc	1211–1188	57
Pri10	gtgatacagattcattgaagaag	1168–1191	cactgttgcaatgtgatggaag	1277–1254	60
Pri11	ctattggaaatgcacttcatgc	1213–1235	cggtggctcatgcctgtaatc	2372–2351	57

[a] Sequence of *MDM2* cDNA clones (Oliner et al., 1992).

per cycle, with a final extension of 72 °C for 10 min after the last cycle.

PCR products were separated by electrophoresis on a 1% low melting temperature agarose gel (NuSieve GTG, Flowgen) and visualised by ethidium bromide staining with UV transillumination. DNA bands amplified from genomic DNA were excised and purified with a QIAquick Gel Extraction kit (Qiagen), by following the procedure recommended by the company.

2.4. Cloning PCR products

Purified PCR products were subcloned directly into the pGEM-T Easy vector (Promega) following the protocol recommended by the company. Ligation products were transformed into *Escherichia coli* JM109 and clones containing the desired inserts were identified by PCR screening. Plasmids were prepared by using the Wizard Plus SV Miniprep system (Promega).

2.5. Sequencing

Sequencing was carried out manually by using PCR product directly as a template or automatically by using plasmid PCR product clone as a template. Manual sequencing was performed using the Sequenase Version 2.0 DNA sequencing system (Amersham, Product No.70770). The automated sequencing was carried out in the central core facility at the University of Newcastle upon Tyne Medical Faculty. The sequences were aligned with the published *MDM2* cDNA sequence, using the DNASTAR sequence analysis software package.

3. Results

3.1. The detection of multiple-sized MDM2 transcripts in human tumours

In our previous studies, we have found five alternative-sized *MDM2* transcripts (*MDM2-a, -b, -c, -d* and *-e*) (Sigalas et al., 1996). Our present investigation of the *MDM2* transcriptional pattern in human bladder tumour samples, but not in normal bladder tissue (data not shown), using RT-PCR has revealed two further transcripts, sized 813 and 707 bp, which we have designated *MDM2-a1* and *-g* (Fig. 1). Sequencing shows that these two transcripts have internal sequence deletions: *MDM2-a1* lacks nucleotides from codons 28 to 222 and codons 275 to 300; *MDM2-g* misses nucleotide codons from 28 to $97^2/_3$ and $114^2/_3$ to 300. Fig. 2 shows the structure of these transcripts in relation to the full-length *MDM2* cDNA sequence and previously described variant transcripts (see GenBank accession numbers AF201370 and AF201371 for details of the sequences).

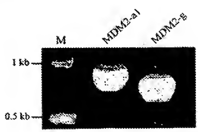


Fig. 1. *MDM2* transcripts were amplified by RT-PCR with nested primers that flank the *MDM2* coding region, as previously described (Sigalas et al., 1996; Matsumoto et al., 1998). M: molecular weight maker.

3.2. MDM2 genomic structure and organisation

We have used long-range PCR amplification, followed by cloning and sequencing to investigate the organisation of the human *MDM2* gene and in particular to define intron–exon boundaries and flanking intronic sequences. Eleven DNA fragments were amplified from genomic DNA with the primer pairs shown in Table 1, which match to the known *MDM2* cDNA sequences. Comparison of the sequences of these PCR products with the published sequences of *MDM2* cDNA clones reveals that the Pri1 primer pair spans introns 1 and 2; while primer pairs Pri2–10 cover one intron per primer pair. However, the region flanked by primer pair Pri11, covering from the 1235th to 2351st nucleotide of the *MDM2* cDNA clone sequence (Oliner et al., 1992), was found not to contain any intronic sequence. Sequence analysis indicates that *MDM2* spans approximately 33 kb of genomic DNA and is separated by 11 introns. Exons range in size from 50 to ≥1161 bp. The size of the introns varies from 121 to ~7000 bp (Table 2). Exon-intron boundary sequences of the 5′ and 3′ splice sites follow the "GT and AG" rule (Table 3; see GenBank accession numbers AF144014 AF144033 for additional intronic sequence data). Table 4 shows the 3′ ends of the intronic sequences adjacent to the intron–exon boundaries, including branch sites and polypyrimidine tracts. The sequences of branch sites have a good match with the consensus sequence YURA*Y (Y: pyrimidine; R: purine; A*: branch point residue). The distances between the branch points and the 3′ splice sites vary from 18 to 111 bp. The C/T content in the polypyrimidine tracts ranges from 53% to 90%.

Comparison of the structure and organisation of the human *MDM2* gene described here with that published for the mouse gene (Jones et al., 1996) indicates that the number of the exons and introns is the same. The size of the coding exons is similar (Table 2). However, the sizes of the noncoding transcribed regions, including exons 1, 2 and 12 and the introns, differ substantially, with the exception of introns 1 and 3.

3.3. Analysis of alternatively spliced variants of MDM2 mRNA

We have detected seven *MDM2* transcript variants (*MDM2-a, -a1, -b, -c, -d, -e* and *-g*) previously and

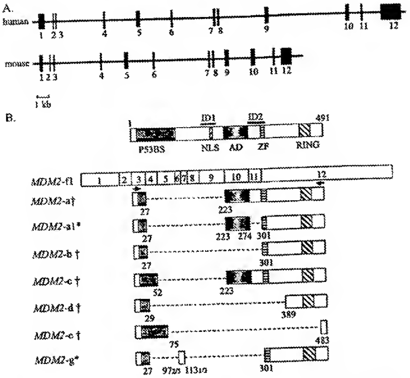


Fig. 2. (A) Genomic organisation of the *MDM2* locus. The top panel presents schematically the human *MDM2* gene structure. For comparison, the mouse *mdm2* gene is illustrated in the bottom panel (Jones et al., 1996). Exons are depicted as vertical bars and numbered underneath. Introns are indicated by horizontal line. (B) The relationship between the various *MDM2* transcripts and the exon–intron boundaries. The second top strip shows the exon organisation of the human *MDM2* gene with its encoded protein above the corresponding exons (top strip). Nucleotide sequence coding for the p53 binding site (p53 BS), nuclear localisation signal (NLS), acidic domain (AD), zinc-finger (ZF), ring-finger (RING) and growth-inhibitory domains 1 and 2 (ID1 and ID2) are depicted with filled boxes. The multiple-sized transcripts that were amplified by RT-PCR with a primer pair (denoted by arrows) are shown with boxes. Dashed lines represent deleted nucleotide sequences. The deletion junctions are numbered by amino acid codons. *detected in this report. †reported by Sigalas et al., 1996.

presently (Sigalas et al., 1996). Some of these variant transcripts have also been demonstrated in astrocytic neoplasms by another group (Matsumoto et al., 1998).

Table 2
Comparison of the human *MDM2* structure with the mouse gene

Exon no.	Exon size (bp)		Intron no.	Intron size (bp)		
	Mouse[a]	Human		Mouse[a]	Human	Difference in size[b] (%)
Exon 1	115	307	Intron 1	~ 500	~ 540	7.4
Exon 2	82	71	Intron 2	160	121	24.4
Exon 3	85	85	Intron 3	~ 4000	~ 4200	5
Exon 4	75	75	Intron 4	~ 2000	~ 3200	37.5
Exon 5	134	134	Intron 5	~ 1750	~ 3500	50
Exon 6	50	50	Intron 6	~ 5100	~ 4000	20
Exon 7	59	68	Intron 7	331	124	62.5
Exon 8	97	97	Intron 8	~ 950	~ 4200	77
Exon 9	164	161	Intron 9	~ 1800	~ 7000	74.3
Exon 10	156	156	Intron 10	~ 1300	~ 900	30.8
Exon 11	78	78	Intron 11	~ 450	~ 2800	83
Exon 12	≥ 615	≥ 1161				

[a] Reference: (Jones et al., 1996).
[b] = bp of difference between mouse and human/bp of the larger corresponding intron.

The alignment of the genomic sequence with the sequences of these seven transcripts discloses that the multiple-sized transcripts resulted from the deletion of multiple entire exons or part of exon sequences (Fig. 2). *MDM2-a* skips exons 4 to 9; *MDM2-a1*, exons 4 to 9 and exon 11; *MDM2-b*, exons 4 to 11; *MDM2-c*, exons 5 to 9; *MDM2-d*, exons 5 to 11, most of exon 4 and the 5′ end of exon 12; *MDM2-e*, exons 6 to 11, the 3′ end of exon 5 and the 5′ end of exon 12. *MDM2-g* splices out exons 4 to 5, and exons 7 to 11 (Fig. 2). Among the skipped exons, exons 4 to 9 and exon 11 were most frequently missed. These commonly excluded exons (with the exception of exon 7) have either less than 60% of C/T content at their upstream polypyrimidine tracts and/or more than 40 bp between their upstream branch points and 3′ sites. Whereas, exons 3, 10 and 12, which are most frequently included in the transcripts, have higher than 60% C/T content within their upstream polypyrimidine tracts and less than 40 bp between their upstream branch points and 3′ splice sites. Since exon 6 contains interrupted codons at both ends, the *MDM2-g* variant, which splices out several exons up- and downstream of

Table 3
Sequence of the human *MDM2* exon/intron boundaries

Exon no.	Sequence of boundary	Exon no.
1	GGAGCAGgtgctggc--intron 1--tttcccagCTGTGTT	2
2	GATCCAGgtaagcac--intron 2--ccttgtagGCAAATG 1	3
3	GACCCTGgttagtat--intron 3--tcttatagGTTAGAC 27 28	4
4	GAAAGAGgtaagctg--intron 4--tatttcagGTTCTTT 52 53	5
5	AGCACAGgtaattct--intron 5--tctacaagGAAAATA 96 98	6
6	CAGCAGGgtaagtta--intron 6--tctctcagAATCATC 113 115	7
7	TCAAAAGgtaatcta--intron 7--atgcttagGACCTTG 136 137	8
8	GAGACAGgtatatat--intron 8--atatccagAAGAAAA 167 169	9
9	GAATCCGgtaatgtt--intron 9--tgtttagGATCTTG 222 223	10
10	TGATGAGgtatatat--intron 10-tttattagGTATATc 274 275	11
11	CTTAGCTgtaagtat--intron 11-cattgaagGACTATT 300 301	12

Exon sequences are shown as uppercase letters. Intron sequences are denoted by lowercase letters. Underlined nucleotides encode amino acids (numbered underneath).

exon 6, has a shifted reading frame after codon 28. The *MDM2-e* variant, which also involves a deletion junction with an interrupted codon sequence, has a shift in the reading frame after codon 484.

4. Discussion

In our previous studies and the results presented here, we detected multiple-sized *MDM2* transcripts in human

Table 4
Branch sites and polypyrimidine tracts of the human *MDM2* gene

Intron no.	Sequences of branch site and polypyrimidine tract (5′ → 3′)	Distance from branch point to 3′ site (nt)	T and C %[a]
1	ctgacttgtctccagctggggctatttaaaccatgcatttcccag/exon 2	>40	59
2	atgatttccagtttcatcgtgtctttttttcctgtag/exon 3	<40	78
3	ttgatggatatgttgctgcagggcctatagttctgggataatttggaagtataatagcagttctttctctttatag/exon 4	>40	53
4	ctaacatttagttatctttaatgctcagaatcatatttgtatttcag/exon 5	>40	60.5
5	ttaatacaaatttattctaaaatgtacatctcttgttatttttttttctgtctacaag/exon 6	>40	68
6	ctgatcctttttctttctctcag/exon 7	<40	90
7	ctaatgtaatattatttgcaaattggaaaggttatttcaacaagttagcttactggttatgttaagtttgttgtattttattttttcctaaatgcttag/exon 8	>40	56.3
8	ttaactttagaactatttattgaaactaagtttctgtacaaataggtactcaaaacagctcaatttctaaatcacagtacagcaattattttttttccttacatatccag/exon 9	>40	57
9	gtgatgtttatcaaattatatatttttttcttgtttag/exon 10	<40	69
10	ctaatgaatgtgttttatag/exon 11	<40	53
11	ctgactgtgtgtcttatttcattgaag/exon 12	<40	61

Branch site sequences are indicated with underlined letters. /: boundary between intron and exon.

[a] The percentage of T and C content in polypyrimidine tracts.

tumour tissues, but not in normal tissues. Detection of alternative spliced forms of *MDM2* mRNAs varied, and appeared to be relatively abundant in most samples. These variants encoded protein products in vitro and were found to transform NIH3T3 cells and to be associated with high-grade and late-stage human cancer (Sigalas et al., 1996 and unpublished data). Our data were supported by the observations of other groups, who reported that alternatively spliced *mdm2* transcripts promoted tumour formation in mouse model (Fridman et al., 2003; Steinman et al., 2004). These collective observations suggest that alternatively spliced forms of *MDM2*, encoding alternative proteins with differing functional capabilities, may play an important role in tumour development. The human *MDM2* genomic map presented here enable us to relate the genomic structure and organisation of *MDM2* to the appearance of variant transcript forms and provide a basis for considering alternative splicing mechanisms.

Pre-mRNA splicing involves precise excision of intron sequences and the ligation of exon sequences. In cases of alternative splicing, the excision may occur at cryptic splice sites; exons may be skipped and introns may be retained. The organisation of the exon–intron boundaries of the human *MDM2* gene indicates that the *MDM2-a*, *-a1*, *-b*, *-c* and -g variant transcript forms (Fig. 2) result from multiple entire exon skipping, because the internal deletion junctions correspond exactly to the location of exon–intron boundaries. However, the deletion junctions of *MDM2-d* and *-e* forms do not correspond to the boundaries between exons and introns, and there are no consensus splicing sequences surrounding them to indicate the possible use of cryptic splice sites. This suggests that they may have resulted from an unusual and possibly aberrant splicing mechanism.

The regulation of alternative splicing involves both *cis* elements and *trans*-acting factors. The *cis* elements include the 5′ and 3′ splice sites, a branch site and a polypyrimidine tract between the branch point and the 3′ splice site. It has been demonstrated that a short distance between the branch point and the 3′ splice site and high C/T content at polypyrimidine tracts give rise to high efficiency of splicing in mammals (Helfman et al., 1988; Libri et al., 1989). Our data show that all the 5′ and 3′ splice sites of the human *MDM2* gene obey the "GT" and "AG" rule. The branch sites also have a good match with the consensus sequence. However, the distances between branch points and 3′ sites, and the percentage C/T content vary between introns. The exons most frequently retained in the splice variants have shorter distances between their upstream branch points and 3′ splice sites and/or a higher percentage C/T content in their upstream polypyrimidine tracts, compared with those exons commonly excluded (Table 2). It suggests that the short distance between the branch point to the 3′ site and the polypyrimidine tract with high C/T content confer high splicing efficiency to their adjacent splicing sites, which is consistent with the observation reported by Smith et al. (1989), Goux-Pelletan et al. (1990), Mueller et al. (1997). However, exon 7 was spliced out with a high frequency although there is only a short distance between its branch point and 3′ splice site and there is also a high C/T content in its upstream polypyrimidine tract. The explanation may be that it is spliced out along with upstream exons. The order of intron removal is governed by preferential binding of splice factors rather than in a sequential numerical order (Lewin, 1994). It may be possible that exon 7 processes splicing before exon 6 does, and the 5′ end of exon 7 may in some circumstances be ligated preferentially to the 3′ end of exon 6. If exon 6 has low splice strength, then in the process of ligation with its upstream exon, it may be outcompeted and skipped out together with exon 7.

Currently, we are not clear why the alternative spliced transcripts appear preferentially in tumour samples, especially in advanced stage and high grade, but not in normal tissues (Sigalas et al., 1996; Bartel et al., 2002). Although variant *MDM2* spliced transcripts have been reported in normal tissues in one study (Bartel et al., 2004), we failed to detect these isoforms in noncancerous tissues. It has been proposed that a mRNA surveillance system exists in cells, which protects them from errors of transcription, mRNA processing, or mRNA transport (Pulak and Anderson, 1993). Mistakes are not uncommon in splicing of RNA from complex genes. Exons can occasionally be skipped (Nigro et al., 1991). In the normal situation, the surveillance system would probably degrade most mRNA with splicing errors as they are transported to the cytoplasm. We speculate that in cancer cells, this system may not function correctly and, consequently, the splice variants may escape degradation. It is also possible that there are mutations in the intron region that cause alternative splicing and the presence of the variants contribute to the cancer. It would be of interest to test this hypothesis by investigating intron nucleotide sequences in tumour samples that show expression of alternatively spliced forms. However, other models can be envisaged; for instance, we cannot rule out the possibility that *trans*-acting factors are involved in the alternative RNA processing by blocking some splice sites and/or enhancing other splice sites.

In conclusion, we have detected two novel *MDM2* alternatively spliced transcripts and have also defined the structure and organisation of the human *MDM2* gene. In addition we have related this information to potential mechanisms by which alternatively sized *MDM2* transcripts are generated. As the alternatively spliced *MDM2* mRNAs have been shown to possess oncogenic potential and to correlate with advanced malignancies to tumour progress (Haines et al., 1994; Sigalas et al., 1996; Steinman et al., in press), our data may assist in clinical diagnosis of sarcomas displaying *MDM2* amplification and alternative splicing of *MDM2* transcripts.

Acknowledgements

We thank Ratchada Suaeyun for her assistance with the work presented in this paper.

References

Bartel, F., Tauber, H., Harris, L., 2002. Altenative and aberrant splicing of MDM2 mRNA in human cancer. Cancer Cell 2, 9–15.

Bartel, F., Pinkert, D., Fiedler, W., Kappler, M., Wurl, P., Schmidt, H., Taubert, H., 2004. Expression of alternatively and aberrantly spliced transcripts of the MDM2 mRNA is not tumor-specific. Int. J. Oncol. 24, 143–151.

Boddy, M.N., Freemont, P.S., Borden, K.L., 1994. The p53-associated protein MDM2 contains a newly characterized zinc-binding domain called the RING finger. Trends Biochem. Sci. 19, 198–199.

Bueso-Ramos, C., Yang, Y., Deleon, E., McCown, P., Stass, S., Albitar, M., 1993. The human MDM-2 oncogene is overexpressed in leukemias. Blood 82, 2617–2623.

Finlay, C.A., 1993. The mdm-2 oncogene can overcome wild-type p53 suppression of transformed cell growth. Mol. Cell Biol. 13, 301–306.

Fridman, J., Hernando, E., Hemann, M., Stanchina, E., Cordon-Dardo, S., Lowe, S., 2003. Tomor promotion by Mdm2 splice variants unable to bind p53. Cancer Res. 63, 5703–5706.

Gaux-Pelletan, M., Libri, D., d'Aubenton-Carafa, Y., Fiszman, M., Brody, E., Marie, J., 1990. In vitro splicing of mutually exclusive exons from the chichen B-tropomyosin gene: role of the branch point location and very long pyrimidine stretch. EMBO J. 9, 241–249.

Haines, D.S., Landers, J.E., Engle, L.J., George, D.L., 1994. Physical and functional interaction between wild-type p53 and mdm2 proteins. Mol. Cell. Biol. 14, 1171–1178.

Haupt, Y., Maya, R., Kazaz, A., Oren, M., 1997. Mdm2 promotes the rapid degradation of p53. Nature 387, 296–299.

Helfman, D.M., Ricci, W.M., Finn, L.A., 1988. Alternative splicing of tropomyosin pre-mRNAs in vitro and in vivo. Genes Dev. 2 (12A), 1627–1638.

Jones, S.N., Ansari-Lari, M.A., Hancock, A.R., Jones, W.J., Gibbs, R.A., Donehower, L.A., Bradley, A., 1996. Genomic organization of the mouse double minute 2 gene. Gene 175, 209–213.

Jones, S.N., Hancock, A.R., Vogel, H., Donehower, L.A., Bradley, A., 1998. Overexpression of mdm2 in mice reveals a p53-independent role for mdm2 in tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. 95, 15608–15612.

Kubbutat, M.H., Jones, S.N., Vousden, K.H., 1997. Regulation of p53 stability by Mdm2. Nature 387, 299–303.

Landers, J.E., Cassel, S.L., George, D.L., 1997. Translational enhancement of mdm2 oncogene expression in human tumor cells containing a stabilized wild-type p53 protein. Cancer Res. 57, 3562–3568.

Leach, F.S., Tokino, T., Meltzer, P., Burrell, M., Oliner, J.D., Smith, S., Hill, D.E., Sidransky, D., Kinzler, K.W., Volgelstein, B., 1993. p53 Mutation and MDM2 amplification in human soft tissue sarcomas. Cancer Res. 53, 2231–2234.

Lewin, B., 1994. The apparatus for nuclear splicing. In: Lwein, B. (Ed.), Genes V. Oxford Univ. Press, Oxford, pp. 914–940.

Libri, D., Marie, J., Brody, E., Fiszman, M.Y., 1989. A subfragment of the beta tropomyosin gene is alternatively spliced when transfected into differentiating muscle cells. Nucleic Acids Res. 17, 6449–6462.

Lundgren, K., Montes de Oca Luna, R., McNeill, Y.B., Emerick, E.P., Spencer, B., Barfield, C.R., Lozano, G., Rosenberg, M.P., Finlay, C.A., 1997. Targeted expression of MDM2 uncouples S phase from mitosis and inhibits mammary gland development independent of p53. Genes Dev. 11, 714–725.

Martin, K., Trouche, D., Hagemeier, C., Sorensen, T.S., La Thangue, N.B., Kouzarides, T., 1995. Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. Nature 375, 691–694.

Matsumoto, R., Tada, M., Nozaki, M., Zhang, C.L., Sawamura, Y., Abe, H., 1998. Short alternative splice transcripts of the mdm2 oncogene correlate to malignancy in human astrocytic neoplasms. Cancer Res. 58, 609–613.

Momand, J., Jung, D., Wilczynski, S., Niland, J., 1998. The MDM2 gene amplification database. Nucleic Acids Res. 26, 3453–3459.

Momand, J., Zambetti, G.P., Olson, D.C., George, D., Levine, A.J., 1992. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell 69, 1237–1245.

Montes de Oca luna, R., Tabor, A.D., Eberspaecher, H., Hulboy, D.L., Worth, L.L., Colman, M.S., Finlay, C.A., Lozano, G., 1996. The organization and expression of the mdm2 gene. Genomics 33, 352–357.

Mueller, A., Odze, R., Jenkins, T.D., Shahsesfaei, A., Nakagawa, H., Inomoto, T., Rustgi, A.K., 1997. A transgenic mouse model with cyclin D1 overexpression results in cell cycle, epidermal growth factor receptor, and p53 abnormalities. Cancer Res. 57, 5542–5549.

Nigro, J.M., Cho, K.R., Fearson, E.R., Kern, S.E., Ruppert, J.M., Oliver, J.D., Kinzler, D.W., Vogelstein, B., 1991. Scrambled exons. Cell 64, 607–613.

Oliner, J.D., Kinzler, K.W., Meltzer, P.S., George, D.L., Vogelstein, B., 1992. Amplification of a gene encoding a p53-associated protein in human sarcomas [see comments]. Nature 358, 80–83.

Oliner, J.D., Pietenpol, J.A., Thiagalingam, S., Gyuris, J., Kinzler, K.W., Vogelstein, B., 1993. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. Nature 362, 857–860.

Pulak, R., Anderson, P., 1993. mRNA surveillance by the *Caenorhabditis elegans smg* genes. Genes Dev. 7, 1885–1897.

Sigalas, I., Calvert, A.H., Anderson, J.J., Neal, D.E., Lunec, J., 1996. Alternatively spliced mdm2 transcripts with loss of p53 binding domain sequences: transforming ability and frequent detection in human cancer. Nat. Med. 2, 912–917.

Smith, C.W.J., Porro, E.B., Patton, J.G., Nadal-Ginard, B., 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. Nature 342, 243–247.

Steinman, H., Burstein, E., Lengner, C., Gosselin, J., Pihan, G., Duckett, C., Jones, S., 2004. An alternative splice form of Mdm2 induces p53-independent cell growth and tumorigenesis. J. Biol. Chem. 279, 4877–4886.

Xiao, Z.X., Chen, J., Levine, A.J., Modjtahedi, N., Xing, J., Sellers, W.R., Livingston, D.M., 1995. Interaction between the retinoblastoma protein and the oncoprotein MDM2. Nature 375, 694–698.

RELATED PROCEEDINGS APPENDIX

None

EXHIBIT 2

GENOME RESEARCH

Is "Junk" DNA Mostly Intron DNA?

Gane Ka-Shu Wong, Douglas A. Passey, Ying-zong Huang, Zhiyong Yang and Jun Yu

Genome Res. 2000 10: 1672-1678
Access the most recent version at doi:10.1101/gr.148900

Is "Junk" DNA Mostly Intron DNA?

Gane Ka-Shu Wong,[1,3] Douglas A. Passey,[1] Ying-zong Huang,[1] Zhiyong Yang,[1] and Jun Yu[1,2]

[1]*Human Genome Center, Department of Medicine, University of Washington, Seattle, Washington 98195, USA;* [2]*Human Genome Center, Institute of Genetics, Chinese Academy of Sciences, Beijing, China*

Among higher eukaryotes, very little of the genome codes for protein. What is in the rest of the genome, or the "junk" DNA, that, in *Homo sapiens*, is estimated to be almost 97% of the genome? Is it possible that much of this "junk" is intron DNA? This is not a question that can be answered just by looking at the published data, even from the finished genomes. One cannot assume that there are no genes in a sequenced region, just because no genes were annotated. We introduce another approach to this problem, based on an analysis of the cDNA-to-genomic alignments, in all of the complete or nearly-complete genomes from the multicellular organisms. Our conclusion is that, in animals but not in plants, most of the "junk" is intron DNA.

Among higher eukaryotes, very little of the genome codes for protein. What is in the rest of the genome, or the "junk" DNA, that, in *Homo sapiens*, is estimated to be almost 97% of the genome? If a region is gene-poor, is that because there are vast deserts of intergenic DNA between adjacent genes, or is that because the few genes that are there are large, with enormous introns?

First, a few definitions are needed. We consider only the euchromatic portion of the genome. The heterochromatic portion (e.g., centromeres and telomeres) is highly repetitive and largely devoid of genes. It is extremely difficult to clone, extremely polymorphic, and unlikely to be sequenced correctly anytime soon. We define the exons and introns as "intragenic" and everything else as "intergenic." This is not to imply that intergenic DNA is nonfunctional, especially as we have incorporated the promoters into our definition. However, promoters are difficult to identify, whereas exons and introns are reliably identified by cDNA-to-genomic alignments. Lastly, we will use the term "genomic length" to indicate the sum of the exons and introns in a given gene and "cDNA length" to indicate the sum of only the exons.

Even after a genome is completely sequenced, it is not a straightforward matter to determine the intergenic fraction. Indeed, any assessment that is based only on the fraction of the genome that has not been identified by the gene annotations is likely to be an overestimate of the underlying reality. Consider how the genes are annotated. Most current procedures (The *Caenorhabditis elegans* Sequencing Consortium 1998; Dunham et al. 1999; Lin et al. 1999; Mayer et al. 1999; Adams et al. 2000; Hattori et al. 2000) employ a combination of EST/cDNA/protein alignments and ab initio exon-prediction programs. Given the incomplete state of the EST/cDNA/protein data, most of the annotated exons are in fact based on the exon-prediction programs, even if parts of certain genes are confirmed by the experimental data. There are two problems (Burset and Guigo 1996; Reese et al. 2000). One is that the exon-prediction programs cannot identify untranslated non-coding exons (i.e., the UTRs). The second, more important, issue is that these programs are not particularly proficient at identifying large genes. There are three reasons: (1) The signal-to-noise ratio can be as low as 1/1000, for the extreme case of a 100-bp exon juxtaposed next to a 100-kb intron; (2) the data sets used to train these programs tend to under-represent the large difficult-to-sequence genes; and (3) the codon-usage statistics, by which the exons are initially identified, are not as informative for the large genes of certain organisms (Wright 1990).

The extent of the large-genes problem is organism dependent. The determinant is the distribution of genomic lengths. If the genomic lengths are distributed over many orders of magnitude, then failure to annotate even a small fraction of the largest genes will leave a much larger fraction of the genome unannotated. In this scenario, there is a critical difference between the following two seemingly similar quantities: the fraction of the genes in the genome that is correctly identified and the fraction of the genome sequence that is labeled as intragenic. The first quantity is far more likely to be correct than the second. It is possible that the total gene count is essentially correct, while, at the same time, the intragenic fraction is significantly underestimated and the intergenic fraction is significantly overestimated. Indeed, this is precisely the problem for the animal genomes.

Our solution is to determine the distribution of genomic lengths entirely from cDNA-to-genomic alignments (i.e., independent of the exon-prediction

[3]Corresponding author.
E-MAIL gksw@u.washington.edu; FAX (206) 685-7344.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.148900.

10:1672–1678 ©2000 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/00 $5.00; www.genome.org

programs). Then, compare the mean genomic length to the mean gene-to-gene distance. The former is taken from the cDNA alignments, but the latter is computed as the ratio of the euchromatic genome size, divided by the gene count, taken from the annotations. Reliable results are expected for *Drosophila melanogaster* and *Caenorhabditis elegans*, because genome sequencing for these organisms is complete and estimates of the gene-to-gene distance are available. For *Arabidopsis thaliana*, the published chromosomes (Lin et al. 1999; Mayer et al. 1999) agree to 4.5%, so we can safely extrapolate to the entire genome. In contrast, for *H. sapiens*, the published chromosomes (Dunham et al. 1999; Hattori et al. 2000) differ by 243%, reflecting the heterogeneity in the gene densities of warm-blooded vertebrates (Bernardi 2000). Coupled with the difficulties of determining the mean genomic length, a result of the lack of large genomic contigs, we refer extensively to the model organism results to guide our interpretations of the *H. sapiens* data.

RESULTS

Figure 1 depicts the distribution of genomic lengths for *H. sapiens*, *D. melanogaster*, *C. elegans*,, and *A. thaliana*. Table 1 is a numerical summary. The animal distributions span 2–3 orders of magnitude, but the plant distribution spans only one order of magnitude. The implication for the large-genes problem can be estimated by considering how many of the largest genes would have to be unidentified for half of the intragenic space to be missing. The figures range from 11% and 10% at one extreme, in *H. sapiens* and *D. melanogaster*, to 30% at the other extreme, in *A. thaliana*. Furthermore, the only organism in which the intergenic fraction is

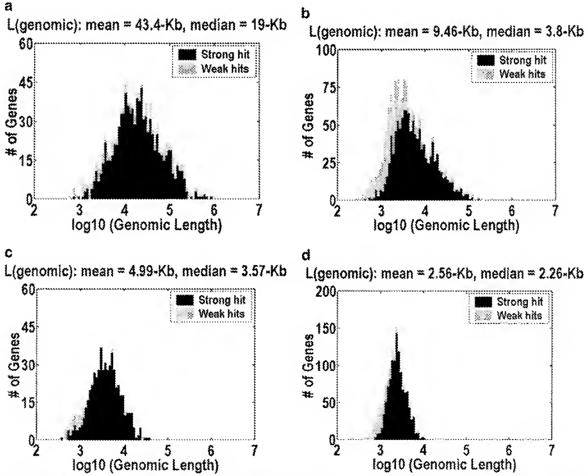


Figure 1 Distribution of genomic lengths for (*a*) *Homo sapiens*, (*b*) *Drosophila melanogaster*, (*c*) *Caenorhabditis elegans*, and (*d*) *Arabidopsis thaliana*. Dark shading indicates strong hits. Weak hits (lightly shaded) represent cDNA-to-genomic alignments with <3 exons or <50% of the cDNA length aligned. An overwhelming majority of these weak hits are actually complete alignments with only one or two exons. Instances in which <50% of the cDNA is aligned represent 7.3%, 3.3%, 1.2%, and 0.9% of the genes in the four organisms, respectively.

Table 1. Estimated Intergenic Fractions

	Homo sapiens	*Drosophila melanogaster*	*Caenorhabolitis elegans*	*Arabidosis thaliana*
Euchromatin	3180000	123000	97800	130000
Sequenced DNA	369000	123000	91000	119000
Gene-to-gene	45.4	9.0	5.3	4.7
cDNA aligned	1061	1628	583	1401
Genomic quality	1.2	23.3	2.4	15.7
Nested genes	6%	8%	4%	1%
05 Percentile	2.5	0.9	0.8	0.9
Genomic length	43.4	9.5	5.0	2.6
95 Percentile	165.5	36.3	14.2	5.4
%, missing half	11%	10%	21%	30%
Intergenic DNA	Discussed in text of article	3%	10%	46%

The first three rows list the euchromatic genome size, the amount of genomic sequence that was analyzed, and the annotation-based estimate of the gene-to-gene distance. The next three rows describe the cDNA alignments. These rows list the number of aligned cDNAs, our quality assessment for the genomic contigs (i.e., the median of the genomic contig size divided by the genomic length for the 95th-percentile gene), and our estimate of the frequency of nested genes (i.e., genes on the reverse strand or inside an intron). The genomic length is given in the next three rows by its arithmetic mean, and its 5th or 95th percentile values. Next, we indicate what fraction of the largest genes would have to be unidentified for half of the intragenic space to be missing. The last row lists the intergenic fraction, computed by correcting the mean genomic length for nested genes, dividing that by the mean gene-to-gene distance, and subtracting the result from one. Note: In *Drosophila melanogaster*, we do not count scaffold joins longer than 1 kb as contiguous when computing the genomic quality. All lengths are reported in kp.

greater than 10% is *A. thaliana*, even though we have included the minor correction for nested genes (genes on the reverse strand or inside an intron). This correction is computed by counting the occurrences of nested genes in our cDNA alignments, and adjusting for the fact that we do not detect every such occurrence because we do not have all of the cDNAs.

The main uncertainty in our method is that we must extrapolate from a subset of the genes to the entire genome to determine the mean genomic length. There will be sampling biases, but they can be categorized and subcategorized as follows: (1) the extent to which cDNA data are enriched for large or small genes, (2) the extent to which genomic data are biased for large or small genes, and then, are the gene-rich regions done first by sequencing projects? Are the contigs large enough for us to align the large genes?

We will argue that the problem is primarily in the genomic data, not the cDNA data. Furthermore, to the extent that there are sampling biases, the tendencies are always to underestimate the mean genomic length and to overestimate the intergenic fraction.

There are two reasons to suspect that biases in the cDNA data will cause us to underestimate the mean genomic length. Keep in mind that large genes are highly correlated with large cDNAs (this paper; data not shown). The first explanation is that full-length cDNAs are extremely difficult to clone, given the ease with which RNA molecules are degraded and the intrinsic bias in the cloning system for smaller inserts. The second reason is that large RNA molecules require more time to transcribe, so large genes might be less highly expressed and more difficult to isolate. However, this expectation is incorrect, because the transcription machinery operates in parallel. As a measure of the expression levels, in *H. sapiens*, we aligned the 1,856,102 ESTs in GenBank against our cDNA data. Multiple reads from the same clone were counted only once. Figure 2 shows that there is no significant variation in EST coverage as a function of genomic length. Notice that the normalization procedures (Hillier et al. 1996) applied to the EST libraries do not affect the rare transcripts, in which we were looking for an effect. The conclusion is that cDNA data, extracted from GenBank, can be representative of all genomic lengths.

Genomic data are biased in two ways. First, there is a sociologic bias toward sequencing gene-rich regions first. Second, even when a genome is complete, lack of long-range contiguity, on the scale of the largest genes, will reduce the estimate of the mean genomic length, because any breaks in the alignment are most likely to occur across the largest introns. Both issues are relevant in the *H. sapiens* data. In Figure 3, we demonstrate that the aligned data are biased toward GC-rich genes, which are of smaller genomic lengths (Bernardi 2000). As for contiguity, we estimate the extent of the problem by computing the ratio of the median genomic contig size to the genomic length of the 95th percentile gene. Ideally, this ratio would be much greater than one. Table 1 shows that it is much greater

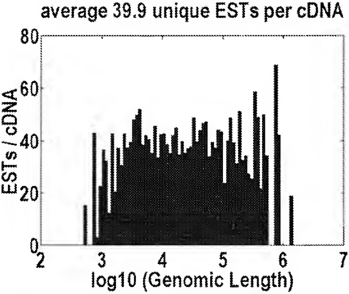


Figure 2 Is the collection of *Homo sapiens* cDNA sequence biased? We aligned the 1,856,102 ESTs in GenBank to our cDNA sequences and plotted the number of aligned ESTs as a function of the genomic length. Multiple reads from the same clone are counted only once. There is no obvious bias, indicating that cDNAs for genes of every genomic length are equally easy to isolate.

than one in *D. melanogaster* and *A. thaliana*. It is only moderately greater than one in *C. elegans*, but that is less important for this organism, because the genomic lengths are not as broadly distributed. However, in *H. sapiens*, the ratio is 1.2, and it would have been even smaller had we not used genomic data from a new division of GenBank in which all of the overlapping clones have been joined together (Jang et al. 1999).

We can estimate the severity of these biases with the different versions of the *D. melanogaster* genomic data. Specifically, we repeated the alignments with the same cDNA data but switched to the 34.9 Mb of finished clone-by-clone genomic data that was available prior to the completion of the whole-genome shotgun (Adams et al. 2000). The contig quality measure is then 2.8, and the resultant mean genomic length of 7.1 kb is off the mark by 34%. By comparing those cDNAs aligned in both data sets, we find that 16% of this effect is attributable to the contiguity problem. The other 18% is attributable to the bias toward sequencing gene-rich regions first. An even more dramatic example of these biases is *Mus musculus*, which has a contig quality measure of 0.3 and a mean genomic length of 9.7 kb. If we assume that there is no difference between *M. musculus* and *H. sapiens*, this estimate is off the mark by 447%. Parenthetically, another unreliable way to estimate the mean genomic length is to extract GenBank annotations. The annotated genes in that 34.9 Mb of genomic data for *D. melanogaster* have a mean genomic length of 3.0 kb, which is off the mark by 317%.

The essential conclusion is that our 43.4 kb figure for the mean genomic length in *H. sapiens* is a substantial underestimate, even if it is already 10 times larger than the training sets used for these exon-prediction programs. However, the gene count itself is also uncertain. The traditional estimate of 70,000 (Antequera and Bird 1993; Fields et al. 1994) has recently been challenged by substantially lower estimates, from 35,000 to 45,000 (Ewing and Green 2000; Hattori et al. 2000; Roest Crollius et al. 2000). How can we interpret the *H. sapiens* data? If we accept the traditional gene count of 70,000, our mean genomic length of 43.4 kb predicts an intergenic fraction of 10%. Suppose we inflate our estimate by the same 34% discrepancy that was observed between the two *D. melanogaster* data sets. The gene count that would be consistent with the same 10% intergenic fraction is

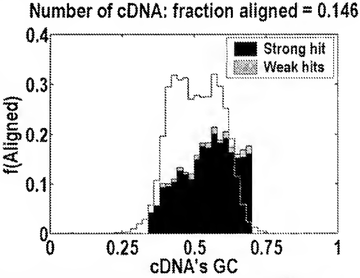
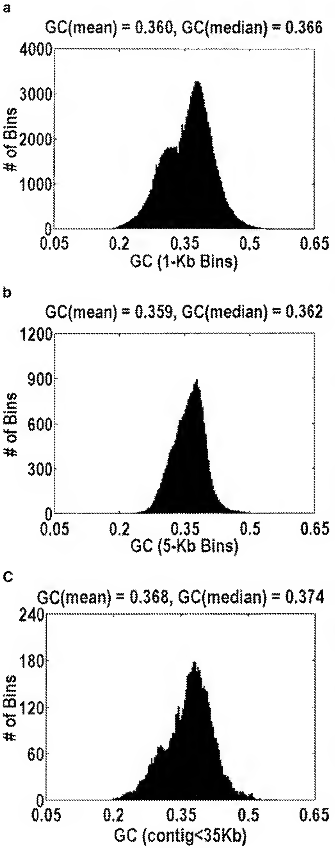


Figure 3 Is the collection of *Homo sapiens* genomic sequence biased? We computed the probability that cDNAs of a particular GC content aligned to genomic seqence, given that only 369 Mb of nonredundant finished genomic sequence were available. The solid line (on an arbitrary scale) indicates the initial collection of cDNAs. The obvious bias toward GC-rich cDNAs is important because these are known to correspond to smaller genes (Bernardi 2000). Dark shading shows strong hits; light shading shows weak hits.



then 51,400. Considering that the contig quality is much worse in *H. sapiens* than in the clone-by-clone *D. melanogaster* data, it is likely that the mean genomic length is underestimated by >34%. Thus, the gene count would have to be substantially less than the current low estimates of 35,000 to 45,000 for our arguments to allow much intergenic DNA.

Given the uncertainty in our method, we cannot give a precise estimate for the intergenic fraction in *H. sapiens*. However, we are prepared to argue that the intergenic fraction in *H. sapiens* cannot be as large as it is for *A. thaliana*, because, at such a high intergenic fraction, the distribution of GC content for genomic DNA is bimodal, as in Figure 4. Fitting the data to a sum of Gaussians reveals that the main mode is centered at 0.382, which is almost identical to the 0.390 GC content of the aligned *A. thaliana* genes. The relative ratio of the two modes implies an intergenic fraction of 30%, which is smaller than the 46% estimate derived from genomic length arguments but not unexpectedly so, because some of the intergenic DNA could have a GC content that is similar to the intragenic DNA. The reason why this bimodality has not been reported previously is that it is extremely sensitive to how the data are plotted. Specifically, the histogram bins must be smaller than the mean genomic length, and smaller genomic contigs (i.e., those sequenced because they contain a likely gene) cannot be used. That said, no such bimodality is observed in *H. sapiens*, *D. melanogaster*, or *C. elegans*, regardless of how the data are plotted.

DISCUSSION

So why do most genome annotation efforts continue to report so much intergenic DNA? One of the most conspicuous features of the recent annotations for *H. sapiens* chromosomes 21 and 22 is the small handful of megabase-sized regions

Figure 4 Distribution of GC content for anonymous genomic sequence in *Arabidopsis thaliana*. The idea that a significant fraction of the genome is intergenic, coupled with the fact that intergenic DNA has a lower GC content than intragenic DNA, suggests that this distribution will be bimodal. However, the bimodality is easily obscured by how the data are plotted. *a* and *b* differ in the size of the bins over which the GC content is computed, 1 kb and 5 kb, respectively. Bin sizes larger than the average gene size of 2.6 kb obscure the effect because every bin is likely to contain a mixture of intragenic and intergenic DNA. *a* and *c* differ in the genomic contigs that are plotted (every contig or only contigs <35 kb, respectively). By removing the large-insert clones favored by the genome centers, what is left behind are those sequences that were analyzed only because they contain a likely gene. Hence, the bimodality disappears.

with absolutely no annotated genes. In all likelihood, each of these regions has one or more large genes, with no counterpart in the EST/cDNA/protein data and which are not being detected by the exon-prediction programs. After accounting for large genes, the remainder of the presently unannotated regions will likely be attributed to untranslated non-coding exons and flanking introns. We must reiterate that the fraction of the genes that is missing does not have to be large to explain away most of the unannotated regions.

What is important is not the precise intergenic fraction or the precise gene count but, at the risk of extrapolating from a limited number of genomes, the differences between plants and animals. There is evidence that plant and animal genomes are organized in different ways. In *H. sapiens*, large genes are caused by a combination of large introns and more introns per gene (this paper; data not shown). At least 35.4% of the total length of the introns in our *H. sapiens* data is due to interspersed repeats (e.g., *Alu* and *L1*). The true fraction is undoubtedly greater, as older repeats, whose sequences are >50% diverged from the ancestral consensus, cannot be identified by existing methods (Smit 1996). Analysis of orthologous genes in *Fugu rubripes* and *H. sapiens* reveals that much of the 10-fold difference in the sizes of these two genomes can be explained by differences in intron sizes (Elgar et al. 1996). In contrast, analysis of syntenic loci among grasses reveals that much of the 40-fold difference in the size of these genomes can be explained by their extensive repeat-filled intergenic regions (SanMiguel et al. 1996; Bennetzen et al. 1998).

The conclusion is that, in animals, most repeats integrate into intron DNA, but, in plants, most repeats integrate into intergenic DNA. Is there something different about the nature of the repeats that insert into animals and plants? Does this dichotomy reflect differences in the operation of the introns and promoters? The answers to these questions will be critical for our understanding of the evolution of large-scale genome features.

METHODS

In *H. sapiens*, cDNA data were extracted from GenBank release 112, but genomic data were downloaded, at the same time, from the new division for nonredundant joined-contigs (Jang et al. 1999). In *D. melanogaster*, cDNA data were taken from release 115 (Dec/15/1999), but genomic data were taken from the whole-genome shotgun (Adams et al. 2000). In *C. elegans* and *A. thaliana*, both cDNA and genomic data were extracted from release 115.

For the cDNA-to-genomic alignments, we required a 98% base pair agreement. We scanned the intron sequences for the consensus splice sites, GT-AG, but we also accepted as a substitute GC-AG, albeit, in <1% of the data. Weak hits, defined as those with <3 exons or <50% of the cDNA length aligned, were plotted separately to verify that they were not anomalous. Immune system-related cDNAs (i.e., with the descriptors immunoglobin, Ig, HLA, MHC, V-region, etc.) were removed. Other redundancies were eliminated, up front by removing all cDNAs that are 90% contained in some other cDNA and post-alignment by comparing the genomic coordinates of the aligned exons. Raw genomic lengths were extrapolated to compensate for incomplete alignments–a small correction even for *H. sapiens*, where a total of 86% of the cDNA lengths was aligned. As another quality control, we required that the exact coordinates of the coding region (i.e., the open reading frame) be known, even though it reduced the number of genes in our final data set.

The partial alignment correction is done by computing an adjusted number of exons, N_{exon}, with a linear extrapolation. The adjusted genomic length, $L_{genomic} = N_{exon}\langle L_{exon}\rangle + (N_{exon} - 1)\langle L_{intron}\rangle$, is extrapolated in a similarly linear manner, with the averages $\langle L_{exon}\rangle$ and $\langle L_{intron}\rangle$ being defined on a per gene basis. Because noncoding terminal exons are generally larger than coding interior exons, both extrapolations are only performed across the coding portion of the cDNA sequence. The intention is to ensure that, if anything, we underestimate the mean genomic length.

ACKNOWLEDGMENTS

We thank Phil Green, Maynard Olson, Carl Tou, and Lee Rowen for many useful discussions and suggestions. This work was partly supported by a grant from the National Institutes of Health (1 RO1 ES09909).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

Antequera, F. and Bird, A.P. 1993. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci.* **90:** 11995–11999.

Bennetzen, J.L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M., et al. 1998. Grass genomes. *Proc. Natl. Acad. Sci.* **95:** 1975–1978.

Bernardi, G. 2000. Isochores and the evolutionary genomics of vertebrates. *Gene* **241:** 3–17.

Burset, M. and Guigo, R. 1996. Evaluation of gene structure prediction programs. *Genomics* **34:** 353–367.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282:** 2012–2018.

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., et al. 1999. The DNA sequence of human chromosome 22. *Nature* **402:** 489–495.

Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B., et al. 1996. Small is beautiful: Comparative genomics with the pufferfish (Fugu rubripes). *Trends Genet.* **12:** 145–150.

Ewing, B. and Green, P. 2000. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* **25:** 232–234.

Fields, C., Adams, M.D., White, O., and Venter, J.C. 1994. How many genes in the human genome? *Nat. Genet.* **7:** 345–346.

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405:** 311–319.

Hillier, L.D., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., et al. 1996. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6:** 807–828

Jang, W., Chen, H.C., Sicotte, H., and Schuler, G.D. 1999. Making

effective use of human genomic sequence data. *Trends Genet.* **15:** 284–286.

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., et al. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402:** 761–768.

Mayer, K., Schuller, C., Wambutt, R., Murphy, G., Volckaert, G., et al. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402:** 769–777.

Reese, M.G., Hartzell, G., Harris, N.L., Ohler, U., Abril, J.F., and Lewis, S.E. 2000. Genome annotation assessment in *Drosophila melanogaster*. *Genome Res.* **10:** 483–501.

Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., et al. 2000. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat. Genet.* **25:** 235–238.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765–768.

Smit, A.F. 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6:** 743–748.

Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* **87:** 23–29.

Received May 23, 2000; accepted in revised form August 29, 2000.